Validation of safety risk prediction tools and traffic models for railway level crossings

A thesis submitted to the University of Huddersfield

in partial fulfilment of the requirements for the

degree of Doctor of Philosophy

Peter Hughes

School of Computing and Engineering, University of Huddersfield

August 2021

Abstract

There are approximately 6000 level crossings in Britain where the trains and road users cross at the same level. In the ten-year period from 2006 to 2016, there were 86 fatalities as a result of collisions between trains and road users at level crossings. Around the world there are a number of safety risk prediction tools in use by road and railway authorities which consider physical and operational features of a level crossing as inputs and produce a prediction of the safety risk for the crossing. There is little information regarding the method of calculation used in any of the tools and no evidence can be found of validation of the results produced by the tools. There is also a large degree of variety between the features that are considered by the tools; the only commonality that can be found is that every tool uses an underlying traffic model *to account for how safety risk varies as road traffic volume increases at level crossings. The most common traffic model is traffic moment – which is the product of road and rail traffic in a day – although some other models are used notably the hypothesis developed by Stott (1987) and the model developed by Peabody and Dimmick.*

Until recently it has not been practical to test the degree to which any of the traffic models correlate with observed collisions due to unavailability of the data. The GB railway infrastructure manager has made information available on the numbers of road users traversing each level crossing, together with the numbers of collisions that have occurred. As such it is now possible to perform more rigorous tests of the degree to which the outputs of traffic models correspond with collisions. Furthermore, in recent years there have been advances in computer technology that have introduced new

techniques to obtain information from observation data; these techniques include machine learning methods that can be used to identify trends and, in many cases, extract meaningful information, from observed data. There is no information in the available literature that shows that either these data, nor these emerging computation techniques have been applied to the study of safety risk prediction tools, which provides a clear avenue for research that is explored in this work.

This work tests:

- *whether it is reasonable to expect safety risk prediction tools to be able to produce reliable estimates of risk;*
- *the degree to which the risk predictions from current safety risk prediction tools correlate with observed rates of collision; and*
- *whether it is possible to use modern data analysis methods to determine a more accurate method of risk prediction.*

The outcomes of this work make a number of contributions to the prior knowledge on level crossing safety, in particular:

- *Whilst safety risk prediction tools are widely used around the world, no evidence can be found of the predictive accuracy of any of the tools.*
- *The various tools are all based on underlying traffic models although, again, there is no evidence of the accuracy of any of the models. Newly available data make it possible to test the models for level crossings in Britain.*
- *When tested, it was found that the most commonly used traffic model – traffic moment – provides a good theoretical model in idealised conditions but does not appear to correlate well with real-world conditions.*
- *In fact none of the traffic models that can be tested were found to correlate well with observed collisions. Remarkably a model based on observation of collisions in the 1930s is better at describing collision rates than a model specifically created in the 1980s to describe British level crossings.*

- *It was found that, whilst none of the traffic models correlates well with observed collisions, there does appear to be a power-law that describes collision rates. Importantly it appears that the rate of collisions per road user decreases as the number of road users increases at a level crossing. This finding is especially significant as it provides the first evidence to support the practice of level crossing closure as a means of improving safety.*
- *A study was undertaken using machine learning techniques to determine whether it was possible to correlate data on physical and operational features of level crossings with rates of collisions. It was found that, as with the previous studies, traffic volumes do correlate to a small degree, however no other correlation can be found in the data.*

Whilst undertaking this work, additional contributions were made, specifically:

- *a meaningful unit of level crossing safety was established, and*
- *a method for comparing observed collision rates against theoretical models that can be used for overdispersed data was identified.*

As well as advancing the theoretical knowledge on level crossing safety, this work provides meaningful results that are useful to the day-to-day management of the railway.

Acknowledgments

In creating this work I am grateful to those who have given their time to support me, in particular my supervisors Prof Coen van Gulijk and Dr Yann Bezin and my reviewer Dr John Stephenson who provided direction and valuable feedback throughout my research. I am also grateful to Dr Ullrika Sahlin of Lund University who assisted me with an understanding of how to perform meaningful statistical analysis of rare events; and Matthew Newall of Huddersfield University who helped hone my understanding of machine learning methods. I am also grateful to Emeritus Professor Andrew Evans of Imperial College London who provided guidance and valuable knowledge on level crossing safety.

I would also like to thank both the Rail Safety and Standards Board (RSSB) and Network Rail for providing data that supported this study. This study relies heavily on historic collision data at British level crossings which is stored in the railway's central database of accidents by RSSB. The provision of these data was a major assistance to me in performing this work. Similarly whilst some data on level crossings is published by Network Rail on their website, there is a large amount of additional data which is not publicly available. The provision of a copy of Network Rail's main database on level crossings provided important extra data that helped ensure the accuracy of the tests performed in this study.

Table of Contents

List of Figures

Glossary

The following terms have a specific meaning within this document.

Term	**Definition**
active warning device	A warning device provided at a level crossing that changes state when a train is approaching or occupying the level crossing.
AHB	automatic half barrier (*q.v.*)
alpha (α) value	The probability that natural variation in a set of data leads to differences so large that the null hypothesis (*q.v.*) is rejected when it is in fact true.
automatic half barrier	An arrangement of active warning devices (*q.v.*) that provides flashing lights and a barrier over the approach carriageway of a level crossing.
descriptive model	A type of traffic model (*q.v.*) that provides a result based on empirical observation rather than mathematical reasoning (see *predictive model*, *q.v.*).
feature	In machine learning (*q.v.*), a property of an object that is used to determine its classification.
label	In machine learning (*q.v.*), a category that has already been applied to data.
machine learning	A method of determining an algorithm based on analysis of data.

Term	**Definition**
null hypothesis	A hypothesis that there is no significant difference between two samples of data: any observed differences are the result of natural variation in the two sets.
overdispersed	A description of a data sample where the standard deviation is greater than the mean value.
overfitting	A statistical phenomenon where a curve has been fitted to data in such a way as to make the residual error between the model and the observations smaller than the natural variation in the data.
passive warning device	A warning device provided at a level crossing that does not change state when a train is approaching or occupying the level crossing (for example a sign).
Peabody Dimmick model	A traffic model (*q.v.*) proposed by Peabody and Dimmick (see United States Department of Transportation, 2007).
predictive accuracy	The degree to which a safety risk prediction tool (*q.v.*) creates results that agree with observed collision rates.
predictive model	A type of traffic model (*q.v.*) that provides a result based on mathematical reasoning rather than empirical observation (see *descriptive model*, *q.v.*).
safety risk prediction tool	A mathematical tool that provides a safety risk estimate for an individual level crossing.
SMIS	Safety Management Intelligence System; a database of safety-related incidents on the GB railway.

Term	**Definition**
SRPT	safety risk prediction tool (*q.v.*)
Stott's hypothesis	A traffic model (*q.v.*) proposed by Stott (1987).
testing set	In machine learning (*q.v.*), a subset of data that is used to test a candidate algorithm.
traffic model	A part of a safety risk prediction tool (*q.v.*) that describes how variation in road traffic volumes affects safety risk.
traffic moment	A type of traffic model (q.v.) that is the product of trains per day and road vehicles per day traversing a level crossing.
training set	In machine learning (*q.v.*), a subset of data that is used to determine a candidate algorithm.
undersaturated	A type of traffic flow where vehicles are moving and are not being delayed by other traffic on the road.
VT	traffic moment (*q.v.*)

Chapter 1: Introduction

1.1 Level crossing safety

The GB railway network has been built up since the 1820s, it currently has 20,000 miles of track and is an essential part of the UK's transport infrastructure operating more than a billion passenger journeys and carrying tens of millions of tonnes of freight each year (Network Rail, 2018). The railway network interacts with other infrastructure, in particular there are more than 6000 level crossings where road users can traverse the rail line. Level crossings are normally classified by their three main characteristics: accessibility, ownership, and types of warning devices. The different types of accessibility of level crossings are those that are on footpaths and can generally be accessed only by pedestrians, compared with those that are on roads and allow for the passage of motor vehicles over the railway. The different types of ownership separate those level crossing that are on public roads and can be accessed by all road users, compared with those that are on private property and are not necessarily available for members of the public. The main distinction in the types of warning devices are those level crossings where there are static signs that warn road users of the presence of a level crossing: such level crossings are usually referred to as having *passive* warning devices. By contrast are those level crossings that have *active* warning devices which are assemblies of flashing lights, bells, and gates which provide a barrier between road users and the railway. Active warning devices operate a short time before the arrival of a train at a level crossing and continue to provide a warning until the train has passed.

Level crossings are distinct from grade separation (bridges or tunnels) that provide physical distance between road users traversing the rail and trains operating on the track. At level crossings there is no such separation: trains and road users cross at the same level and there remains a possibility for a collision to occur.

When collisions occur, there is the potential for serious consequences, unfortunately, in the ten-year period from 2006 to 2016, there were 86 fatalities as a result of collisions at British level crossings (RSSB, 2016). Reducing the number of collisions and fatalities is a high priority for the railway (RSSB, 2016), however the provision of warning devices at level crossings is expensive, and the cost of grade separation can be much higher (Wullems *et al.*, 2013). To date, Network Rail has spent £200 million on improvements to level crossings (Global Rail Review, 2018).

1.2 Level crossing safety risk prediction tools

Given the high cost of interventions to improve safety at level crossings, railways around the world typically maintain lists of planned works they intend to perform in the future. In determining which interventions have the highest priority, railways often use *safety risk prediction tools* (SRPTs) which provide a method to compare the safety risk amongst level crossings. To calculate a risk score, these tools usually consider a combination of physical and operational characteristics of a level crossing *such as*: the number of road approaches to a level crossing; the distance on the approach that a road user can see the level crossing; or the number and speed of trains traversing the level crossing. The calculated risk scores are used to prioritise safety investment for new safety interventions, *for example*: improving sighting distances for road users approaching a level crossing; installing automatic warning devices; or even replacing a level crossing with a grade separated crossing (Office of Rail and Road, 2011). The Rail Safety and Standards Board (RSSB, 2007) state that SPRTs:

> *allow risk assessments to be carried out without having to conduct a bespoke risk assessment for each crossing, although bespoke data may be used, thus permitting assessments to be done with a higher degree of consistency and reduced effort. A model that produces a quantitative measure of risk (either a relative 'score' or an 'actual' measure of risk such as fatalities per year) allows the identification of highest risk crossings which can then be regarded as highest priorities for taking action to reduce risk.*

The SPRTs in use by railways around the world vary in the features they use as inputs and their methods of calculation, however the method of calculation is not always published. It can therefore be expected that different tools would provide different risk predictions for a given level crossing. There is no clear evidence of validation of any of the SRPTs, and it is not clear to what degree the risk predictions correspond with observed collision rates.

Despite the difference between the differed SRPTs used by different railways around the world, a common feature is that each tool contains some form of *traffic model* which describes how collision rates are expected to vary with changes in road and rail traffic volumes (RSSB, 2007). When considering the effect that road and rail traffic volumes have on rates of collisions, it is axiomatic that a collision can occur only when the level crossing is occupied by a road vehicle and a train at the same time. The simplest model of how risk increases with traffic is to assume that if the road traffic over a level crossing were to double then there would be twice as many opportunities for a collision to occur; similarly a doubling of train volume would also double the opportunities for collision. If collisions occur at random with the arrival of each road vehicle then the number of collisions at a level crossing would be proportional to the product of the number of road vehicles and trains using a level crossing in a given time (Hughes, 2002). The product of road vehicles and trains traversing a level crossing is known as the *traffic moment* (Evans and Hughes, 2019) and often referred to as *VT* being the product of the number of road vehicles per day (*V*) and the number of trains (*T*). Many SRPTs assume that the rate of collisions at a level crossing is directly proportional to the traffic moment. As such, traffic moment is considered a normaliser for level crossing collisions: to compare the relative safety risk of two level crossings it would necessary to divide the observed collisions in a given period by the traffic moment.

Whilst the majority of the SRPTs use traffic moment as the normaliser for collision there are a few tools that use other methods of determining how road traffic volumes affect collision rates. As with the SRPTs overall, there is currently no validation of the rate at which collisions vary at level crossings with varying road traffic volumes.

Whilst the majority of SRPTs use traffic moment as a normaliser, it is notable that this is not the case for *all* SRPTs, in particular there are two other traffic models in use, *Stott's hypothesis* and the *Peabody Dimmick* model. Where these models are used, they are applied in place of traffic moment as a normaliser.

Until recently it has not been practical to test the degree to which any of the traffic models correlate with observed collisions due to unavailability of the data. The GB railway infrastructure manager, Network Rail, has published data on the numbers of road users traversing each level crossing, together with the numbers of collisions that have occurred (Network Rail, 2017). As such it is now possible to perform rigorous tests of the degree to which the outputs of traffic models correspond with collisions. Furthermore, in recent years there have been advances in computer technology that have introduced new techniques to obtain information from observation data (Golio, 2015), these techniques include *machine learning* methods that can be used to identify trends and, in many cases, extract meaningful information, from observed data. There is no information in the available literature showing that either these data, nor these emerging computation techniques have been applied to the study of SRPTs, which provides a clear avenue for research that is explored in this work.

1.3 Research method and intended contribution of this work

Traffic moment is widely used as a normaliser in SRPTs, however there is a lack of evidence to support the validity of this approach; there is a need to confirm whether it is appropriate to use traffic moment in this way. Such a test would be relatively simple to carry out since the method of calculation for traffic moment is clear (it is simply the product of road and rail traffic volumes in a day), unlike the full methods of calculation used in SRPTs which are not publicly available and, therefore, cannot be tested. Furthermore, the methods of calculation for both Stott's hypothesis and the Peabody Dimmick model are published. Using the newly available data on road and rail traffic volumes, it would therefore be possible to test the degree to which these traffic models correlate with observed collisions at level crossings.

It would be desirable to test the correlation between the results of SRPTs – rather than just their traffic models – with observed collisions. It can be expected that the risk predictions made by SRPTs are not exactly the same as the results of the calculation of traffic models, instead it can be expected that the risk prediction is affected by other factors such as road traffic speed, or the number of operational railway tracks. However the fact that all SRPTs contain an underlying traffic model indicates that traffic models are considered important in the calculation of safety risk. Furthermore it can be expected that over a large enough sample, the overall effect of traffic models on risk prediction can be tested.

Furthermore, using publicly available data that regarding level crossing characteristics, together with modern advanced techniques for identifying patterns in data, it is possible to perform tests to determine whether there are any correlations between the characteristics of level crossings that are reported in the available data and observed rates of collisions.

Given the serious consequences that can result from collisions at level crossings, and the expense involved in providing warning devices, it is in the public interest to have confidence in the methods that are used to allocate funding for level crossing safety interventions. However since there is information regarding the various traffic models that underpin the tools has been made available, it is possible to test the way in which road and rail traffic volumes affect collision rates. In 2017, Network Rail published data on road traffic volumes at level crossing traffic, which had not previously been available. These data provide an opportunity for an empirical study to determine whether, and to what degree, road traffic volumes affect collision rates at level crossings.

The purpose of this research is to use the newly available data together with the information that is available on the SRPTs to establish:

- whether it is reasonable to expect SRPTs to be able to produce reliable estimates of risk;
- the degree to which the risk predictions from current SRPTs correlate with observed rates of collision; and
- whether it is possible to use modern data analysis methods to determine a more accurate method of risk prediction.

A novel aspect of this work is that it brings data from the railway in a way that has not previously been performed. Specifically the study uses Network Rail data regarding level crossings and road traffic volumes and data regarding collisions at level crossings. The study considers two sources of collision data: collision data from Network Rail, as well as data from the database of all railway accidents in Britain. The combination of these data sources is a significant novelty of this study. Comparing the observed collision rates, normalised by road traffic volumes, with the traffic models is a further novelty of this study.

It is intended that this study will not only provide new information in the study of level crossing safety, but it will also provide information that is genuinely useful to the GB railway. An improved understanding of the correlation between road traffic volumes and collision rates at level crossings can better inform safety management strategies of the railway. Using advanced computational methods a test will be performed to determine whether it is possible to develop a method of calculation that uses the available data on level crossing characteristics to create an accurate predictor of collision rates. In effect this work would develop a validated SRPT for level crossings. It is intended that this study will contribute to the on-going efforts to improve the safety of the railway.

1.4 Scope of this research

This research considers collisions between road vehicles and trains traversing level crossings. In this study, level crossings are classified in accordance with the scheme used by Network Rail (2017) and Evans and Hughes (2019). There are three main classes of level crossing: railway-controlled; those with automatic warning devices; and those passive warning devices (refer to Section 2.3.1 for a fuller description of these classes). Each class is further divided in two sub-classes: those level crossings that are accessible on public roads; and those that are accessible only from private property or accessible only to railway staff. The scope of this study includes all classes of level crossing. However the level crossings are not evenly distributed between the classes: there are approximately 40 times as many passive, private level crossings are there are railway-controlled private level crossings. For the classes where there are few level crossings, there are correspondingly fewer recorded collisions. Nevertheless data is available on all classes of level crossing, and for the sake of completeness, all classes have been included in the scope of this study.

Within the scope of this study, a collision is considered to be any event where a train and a road vehicle come into contact on a level crossing regardless of whether a train struck a road user who is already on the level crossing panel, or whether a road user struck a train. This research does not consider other types of accident that may occur at a level crossing; such as derailments or train-to-train collisions, that are not the result of collisions with vehicular road users at level crossings. Also beyond the scope of this work is any analysis of collisions between pedestrian road users and trains: there are fundamental differences between pedestrian movements and the operation of road vehicles. For example road vehicles are more limited in how they can move: vehicles in a queue over a level crossing cannot disperse in the way that a queue of pedestrians would be able to. Furthermore it takes considerably more distance to bring a road vehicle to a stand when it is moving at speed than it would for a pedestrian to stop on the approach to a level crossing. The study also does not include collisions with equestrians, users of mobility aids including mobility scooters, nor user of toys such as skateboards.

The study also does not consider collisions where it is believed that the collisions occurred as a result of a motivation by the road user to self-harm (suspected suicide events) since the purpose of the study is to understand the managerial controls that can be put in place to reduce the numbers of collisions. The controls that exist at level crossings are mostly visual and audible warning devices which can be disregarded by a person who is motivated to purposefully collide, indeed advanced warning of an approaching train actually provides information that a road user needs if they are intended to purposefully cause a collision. Enforcement controls, such as red-light cameras (which detect and photograph vehicles that traverse the level crossing whilst stop signals are showing), allow for post-hoc punishment of people who contravene rules. Deterrents based on future punishment can be expected to have reduced impact on a person who is intending to avoid the future.

The data for this study have been obtained from GB railway authorities, specifically Network Rail and RSSB, consequently the scope of this study covers only level crossings in Britain that are operated by Network Rail. Since the method of calculation of the existing SRPT is not known, it is not possible to propose modifications to the existing tool. However, where possible, the study intends to identify a model, or models, that provide better correlation with the observed collisions than the existing traffic models. This study will look to identify if correlations can be found between particular warning devices, or combinations of warning devices. However the work will not look to determine the cost-benefit of specific devices, since new technology is continuing to make new types of warning device available that, in many cases, are substantially less expensive than existing devices (Wullems *et al.*, 2013).

1.5 Structure of this thesis

Following this introduction, Chapter 2 reviews prior work that has been undertaken that is relevant to this research. The review considers a range of literature including those that address the theory of collision causation at level crossings, the types of warning device, safety risk prediction tools, and the traffic models used in the tools. The review identified a number of traffic models with *traffic moment* being the most widely used. The review concludes by examining the literature on emerging methods for determining data-driven safety risk prediction tools.

Based on the findings of the review, the study in Chapter 3 applies two approaches to test the validity of traffic moment as a normaliser for collisions at level crossings. The approaches are mathematical derivation and Monte Carlo simulation. The purpose of this study is to use a theoretical method to test the simplest, most common, element of many SRPTs to test whether it is reasonable to expect SRPTs to be able to produce reliable estimates of risk.

To test the degree to which the risk predictions from current SRPTs correlate with observed rates of collision, Chapter 4 describes the experimental method that was used including the method to prepare the source data as well as the method of calculation to be applied given the overdispersed nature of the data. The results of the experiment are presented in Chapter 5, which also provides a discussion of the interpretation of the results. Chapter 6 describes a study to determine traffic models based on empirical study of the data and discusses the implications of the results with particular application to whether the current approach to close level crossings to improve safety can be supported by the available evidence.

Chapter 7 describes a study that was undertaken to establish whether it is possible to use modern data analysis methods to determine a more accurate method of risk prediction. The findings of the overall study are summarised in Chapter 8 which provides a discussion of the implications of the results and ties together some of the findings of the literature review in Chapter 2 to provide recommendations that employ emerging technologies. A conclusion is provided in Chapter 9.

1.6 Contribution

The following contribution to current knowledge has been made in this chapter:

Contribution 1: It has been identified that there is a gap in the knowledge of SRPTs for level crossing and that there is an opportunity to advance the current state of knowledge by using newly available sources of data and by combining data sources on level crossings and observed collisions in a way that has not previously been performed.

Chapter 2: Background to level crossing safety risk prediction

This chapter provides background to the key concepts considered in this work. Firstly there is a review of the literature on level crossings and the relationships between physical characteristics of level crossings and collisions rates. There is also a consideration of the warning devices that are used at level crossings as well as a consideration of other methods of reducing collision rates at level crossings.

This chapter also includes a discussion of safety risk prediction tools that are used for level crossing risk assessment and the underlying traffic models that are used for these tools as well as considering emerging approaches to developing data-driven safety risk models for the railway.

2.1 Safety risk

Risk is defined by the International Organization for Standardization (2018) as "*the effect of uncertainty on objectives*", whilst this broad definition of risk applies to all effects on objectives – whether these effects are desirable or otherwise – it is common that structured risk management activities concentrate on only the adversely effects on objectives. Risk management activities usually measure risk in terms of the likelihood of occurrence of a specific impact, or type of impact, on objectives. When considering the management of *safety* risks, the types of impact are often categorised as the types of injuries that may result from the occurrence of an uncertain event: for example minor injuries, major injuries or fatalities. The likelihoods of specific impacts are similarly categorised to describe those outcomes that are expected to occur frequently and those that are expected only rarely. Higher risks are those that have the largest impact on objects and those that are expected to occur more frequently. Figure 2.1 has been adapted from Jordan *et al*. (2018) and shows how a matrix is used to classify risks in three categories: *low*, *moderate*, and *high*.

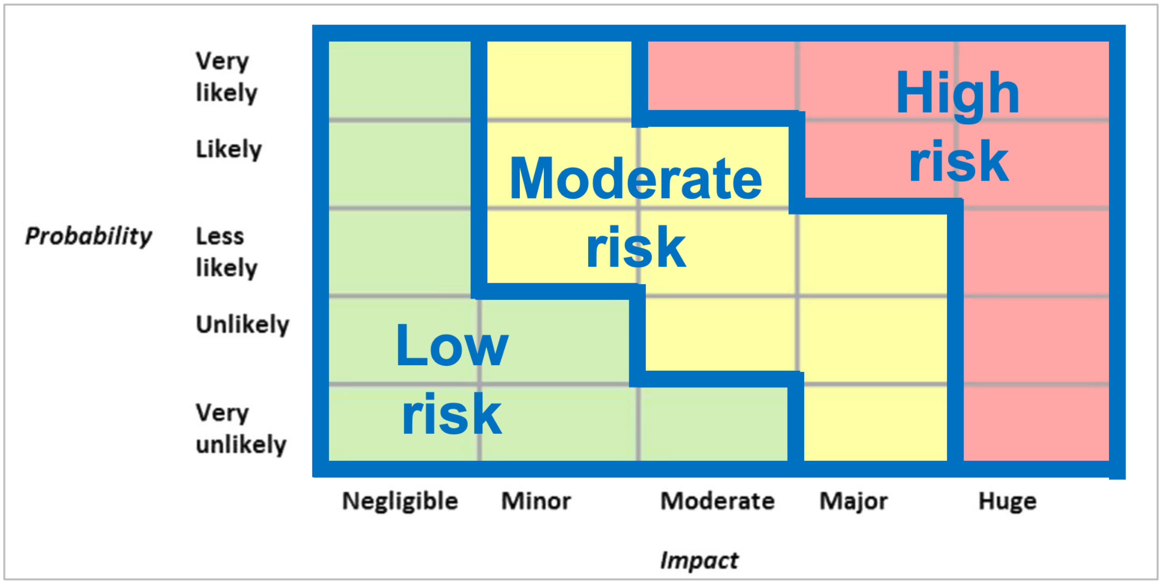


Figure 2.1: Example matrix for risk classification, adapted from Jordan *et al.* (2018)

The main objective of a level crossing is to allow road and rail traffic to traverse the same area without collision and with minimal delay to road users and trains. In this situation the risk arises as a result of the uncertain behaviour of road users: whether they will observe, understand, and correctly comply with the requirements to yield to trains and, therefore, avoid collisions.

Collisions at level crossings can result in a range of impacts to the health and safety of individuals. In many cases collisions result in fatalities to road users and perhaps also to train crew and passengers on trains. In other cases collisions can result in damage to property but no major injuries: for example such an outcome can occur if a road vehicle is struck by a slow-moving train that causes damage to the exterior of the vehicle but does not cause physical injury to its occupants.

The specific outcomes that result from a collision between a train and road vehicle can vary depending on a number of factors such as the degree of physical protection provided by a vehicle, safety devices within the vehicle, or even pre-existing health conditions of people in a vehicle. Currently there is no theory to describe the specific safety impacts that can be expected as a result of collisions at level crossings and therefore these factors are considered to be random since they are beyond the control of road and rail authorities. For this study it is considered that the key objective of road and rail authorities is to avoid all collisions between road users and trains. As such the only measure of impact in this study will be whether or not a collision results from a road user traversing a level crossing; the measure of risk will be affected solely by the likelihood of collisions occurring.

2.2 Theory of level crossing collision causation

Much of the literature on level crossing safety look to establish correlations between physical and operational characteristics of level crossings and collision rates. Although not explicit in any of the literature there appears to be an underlying assumption that the physical and operational characteristics of a level crossing affect road users' situational awareness and motivation to stop at a level crossing, which in turn affects the likelihood of a collision. This model of causation is shown diagrammatically in Figure 2.2.

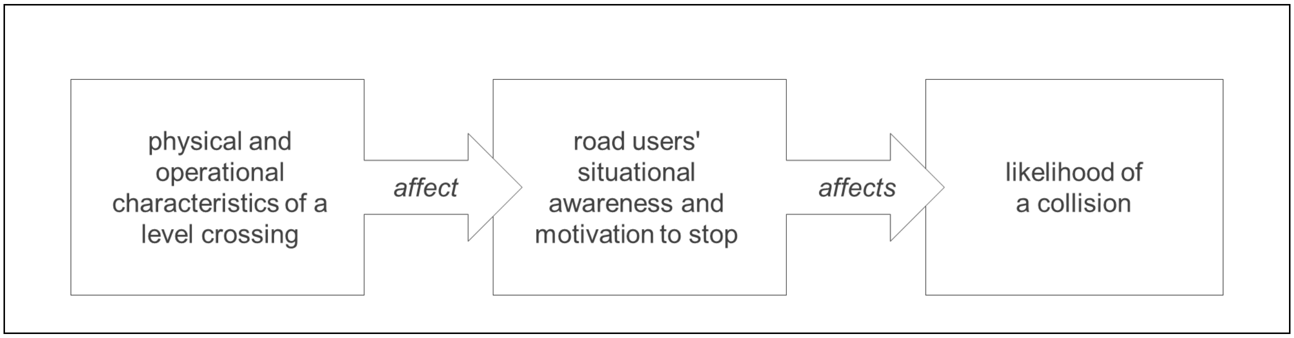


Figure 2.2: Diagrammatic representation of assumed causation model

The assumption that physical and operational factors affect collision rates can be found in much of the literature, for example, Oh *et al.* (2006) who studied behaviour at level crossings in Korea and concluded that "*the proximity of crossings to commercial areas …*[is] *associated with larger numbers of accidents*". Larue (2016) found that the longer the time between the start of flashing light warnings at the arrival of a train at a level crossing the greater the likelihood of a collision. Starčević *et al.* (2016) found that rumble strips on the approach to a level crossing correlate with fewer collisions. Haleem (2016) found that the likelihood of fatalities at private level crossings was influenced by the train speed.

The main observation on reading the literature on level crossing safety is that whilst many researchers have performed many studies and have discovered specific correlations between physical and operational features of level crossing and the occurrence of collisions there appears to be no overarching consensus amongst the researchers. The researchers do not attempt to place their discoveries within a risk framework that describes level crossing safety in general, nor can any such overarching framework be found in the literature. To date, the many studies that have been carried out on level crossing do not contribute to a general theory.

2.3 Classes of level crossings and the hierarchy of controls

The one clear factor that has been shown to affect safety risk at level crossing is level crossings is the nature of the warning devices. Evans and Hughes (2019) demonstrate that the rate of fatalities per road user traverse reduces by at least an order of magnitude at level crossing where there are active warning devices. This reduction in fatality rate occurs for both vehicular level crossings as well as pedestrian level crossings, despite the large underlying differences in rates of collision.

2.3.1 Passive and active warning devices

Passive warning devices are fixed signs that mark the presence of a level crossing and indicate to road users that there is a need to check for the presence of trains before traversing. These warning devices are referred to as passive warning devices since the warning provided by the signs does not change state to indicate the presence of a train.

Conversely, active warning devices are signs that change state – often by showing flashing red lights and sounding bells or alarms – to indicate that a train is approaching or is occupying the level crossing. Active warning devices sometimes include provision of a barrier across either all road carriageways or, in some cases, across only the approach carriageway to the level crossing. In Britain, wherever a barrier is provided across all carriageways, the operation of the warning is controlled manually by a railway employee who can observe the level crossing either directly, or by close-circuit television (CCTV). Where the barrier is across only the approach carriageway, the operation of the warning is controlled automatically by the approach of a train, this type of level crossing is therefore known as an *automatic half-barrier* (AHB) level crossing. Figure 2.3 which is reproduced from the Office of Rail Regulation (2011) provides schematic view of an AHB level crossing.

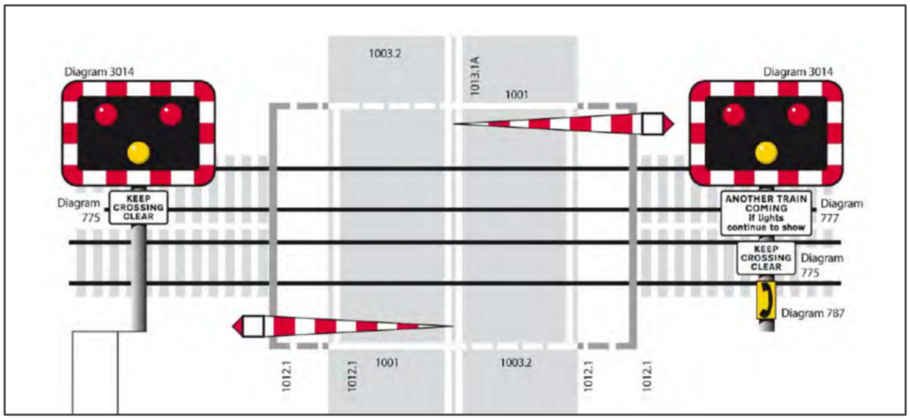


Figure 2.3: Schematic layout of an automatic half-barrier (AHB) level crossing reproduced from the Office of Rail Regulation (2011)

In Britain the minimum time between the warning starting and a train arriving at a level crossing is 13 seconds (Office of Rail Regulation, 2011, §2.48). Whilst this is the minimum time, the actual warning time may be longer especially when the warning is manually activated by a railway employee (RSSB, 2019).

2.3.2 Hierarchy of warning devices

The one area within the literature where there does appear to be consensus is in the assumption that active warning devices contribute to lower safety risk than passive devices, and furthermore that warning devices with barriers lead to a lower safety risk than those without barriers. It is particularly noteworthy that this belief in the relative safety risk arising from each category of warning device has persisted for a long time, for example the model of level crossing collisions developed by Peabody and Dimmick in

1941 (US DoT, 2007) gives a lower prediction of collision rates if active warning devices are present at a level crossing compared with only passive devices. The term *hierarchy* is often used to describe the way in which it is believed that some warning devices are superior to others in reducing safety risk, for example ALCAM (2007) and Wullems *et al.* (2013). Despite the belief in a hierarchy of controls being ubiquitous amongst level crossing safety practitioners for such a long time there is no consensus on the exact categorisation of warning devices. It is common in the literature for the hierarchy to be expressed as being comprised on only three types of device, *viz.*:

- active warning devices with barriers;
- active warning devices without barriers; then
- passive warning devices.

However Baker and Heavisides (2007) describe the hierarchy as:

a. manually controlled level crossings;
b. automatic level crossings with half barriers;
c. automatic level crossings with no barriers;
d. passive level crossings with gates; then
e. passive level crossings with no gates.

A further problem with the concept of the hierarchy of warning devices is the lack of evidence to support its correctness. For a long time there remained little empirical data to show that active devices are actually better than passive devices at reducing safety risk. Hughes (2012) highlights that within Australia there are more fatalities at level crossings with active warning devices than at level crossings with passive devices. However the paper discusses that since active devices are generally installed at level crossings with larger volumes of road and rail traffic, these are the level crossings where there are more opportunities for collisions. Therefore, when normalised by road vehicle traverses, it is not clear whether active warning devices do reduce risk in accordance with the hierarchy. Hughes sums up the discussion by stating "t*he available evidence in Australia is mute on whether active warning devices provide any better protection* [than passive warning devices]".

More recently, however, Evans and Hughes (2019) note that "*data has been made available by the GB railway infrastructure manager, Network Rail which provides a detailed inventory of level crossings in GB. The data include the type of each crossing, and the numbers of crossings or 'traverses' per day*". These data allowed a study of fatalities at British level crossing for a number of categories of warning devices and normalise the fatalities by road user traverses. The results of the study are the first empirical results to support the long-standing belief in the hierarchy, however in their study they considered only three types of vehicular level crossing, *viz.*:

- railway-controlled (which corresponds with category a. listed by Baker and Heavisides above);
- automatic (corresponding to category b.); and
- passive (corresponding to category e.).

Within this three-level hierarchy, their results showed that per traverse each type of level crossing had approximately an order of magnitude fewer fatalities than the type below it.

2.3.3 Other warning devices and the three Es

Aside from the provision of signs, lights, bells and barriers, there are other devices that are used to warn road users of the presence of level crossings. The technical manual for the Australian SRPT (ALCAM, 2007) list a number of other devices including:

- overhead mounted (mast arm) traffic control;
- passive tactile advance warning (*e.g.* rumble strips);
- Rail-X pavement marking;
- hand signallers (also known as *flagmen*);
- street lighting at crossing; and
- maintenance program for vegetation.

Whilst these are common warning devices at level crossings, none of the literature discussing a hierarchy of warning devices makes any mention of these devices.

Hughes (2002) introduces another categorisation of risk control at level crossings, the *Three Es*: *engineering*, *education* and *enforcement*. *Engineering* refers to physical controls at, or near to, the site of a level crossing that provide either passive or active warning of the presence of a level crossing and the need for road users to take action. *Education* refers to public information and education programmes aimed to promote an understanding of the safety risk associated with level crossing and encourage correct behaviour by road users. Figure 2.4 shows an example poster used in an education programme conducted in Britain.

Figure 2.4: An example poster used in an education programme in Britain

Enforcement programmes are comprised of a method of detecting then penalising road users who contravene the requirement to yield to trains at level crossing. In practice such enforcement is usually applied only at level crossings with active warning devices where there is a clearly defined period when road users must keep clear of the level crossing. During a study at a level crossing in Croatia, Barić *et al.* (2018) noted that on the day when a uniformed police officer was present "*the proportion of illegal crossings by pedestrians and cyclists alike fell nearly to zero*". However it is not clear that the benefits of either education or enforcement programmed extend after the programmes have finished. Whilst this is only a small piece of evidence, it is clear that enforcement programmes may be a practical measure to reduced collisions at level crossings – at least during periods when enforcement agents are visible to road users.

Despite education and enforcement controls being widely used in an attempt to reduce level crossing safety risk, the discussions of a *hierarchy* of controls never refers to these two classes of control. Again it appears that whilst there are many researchers and railway safety agencies working to reduce safety risk at level crossings, there is no overarching theory of that addresses the entire scope of risk management at level crossings.

2.4 Safety risk prediction tools

Provision of warning devices at level crossings can be expensive (Wullems *et al.*, 2013). In order to ensure that the benefit from the expenditure on warning devices can be maximised, road and railway authorities around the world attempt to determine the safety risk at individual level crossings and then prioritise the provision of controls using a risk-based approach to provide optimal risk reduction for a given cost. In doing so, road and railway authorities around the world have adopted a range of SRPTs (RSSB, 2007).

The widespread use of SRPTs raises a number of interesting points. Firstly is that the purpose of an SRPT is to attempt to determine some measure of safety risk at individual level crossings. The tools vary between:

- those that attempt to create an *absolute* rating of risk in that they attempt to predict the number of collisions that may occur during some future period; and
- those that create a *relative* measure of risk in that they rank level crossings in order of safety risk without making predictions about numbers of future collisions.

In either case, an underlying assumption in such tools is that it is in some way possible to determine safety risk at level crossings. This assumption conflicts with the fact that, as discussed, there does not appear to be any overarching theory of the underlying factors that affect safety risk at level crossings. Therefore it is not clear how an accurate SRPT can be created. Regardless of this lack of an underlying theory such tools abound, RSSB (2007) identified 23 SRPTs in use around the world.

A second observations is that the various SRPTs use different methods of calculation to determine safety risk. RSSB, note that obtaining information about each of the tools was a difficult task and, in some cases, the information was not available. In other cases, information was shared confidentially with RSSB for the purposes of their study and is not generally available. The information that is available, especially from RSSB's detailed report, makes clear that different methods of calculation are used to estimate safety risk in the different tools. This use of different methods of calculation raises a question. It is likely that the different methods of calculation would each lead to different risk estimates for a given level crossing. It can therefore be expected that some of the SRPTs produce results that are more accurate than others. In fact it is possible – perhaps likely – that some SRPTs produced inaccurate risk estimates when compared with observed collision rates.

It is theoretically possible that each SRPT is correct within its own domain. Perhaps the differences in road and railway operations leads to innate differences in level crossing safety risk that require different SRPTs. Whilst not impossible, such a prospect seems infeasible. All over the world motor vehicles are fundamentally similar in design, furthermore the designs of warning devices around the world are uncannily similar:

overwhelming most warning devices uses signs that show a steam train silhouette; active warning devices are, overwhelmingly, flashing red lights accompanied with bells or sirens. The proposal that different jurisdictions have fundamentally different causes of collisions at level crossings does not accord with intuition, neither is there any evidence in the scientific literature that this is the case. Again, there is no overarching theory of level crossing safety that would allow a reasoned discussion of how safety risk can vary as a result of the different road and railway operations in different jurisdictions.

The most important observation, however, is that no evidence can be found of any tests having been carried out to determine the degree to which the predictions of any SRPTs correlates with observed collisions. This absence of evidence is an serious impediment to the development of level crossing safety theory. Since the SRPTs are often used by public road and railway authorities as methods of prioritising public spending, it might be considered that the methods of calculation and the results of tests of predictive accuracy of the models would be in the public interest and the absence of tests could be a concern to the public. Cynically it could be imagined that tests might have been conducted but that the correlation between prediction and observation was too poor for the road or rail authorities to feel confident publishing the results. If this were the case, then there is an even greater cause for concern.

2.5 RSSB review of safety risk prediction tools

In 2007 RSSB published a report titled *Use of Risk Models and Risk Assessments for Level Crossings by Other Railways* which describes a review of 23 tools for determining safety risk from thirteen countries. In undertaking this work RSSB note:

Significant effort has been made to identify all level crossing models in use or being developed in rail administrations around the world. It is, however, impossible to be certain that the above list includes all models that exist or that are undergoing development.

The following sections provide an overview of the risk prediction tools reviewed by RSSB and the use of traffic models in the tools.

2.5.1 Overview of tools

In their review, RSSB provide a general description of SRPTs as tools that "*tools that allow risk assessments to be carried out without having to conduct a bespoke risk assessment for each crossing, although bespoke data may be used*". Whilst their review does not provide a detailed description of the algorithm used in any of the tools, several examples are provided of the physical and operational characteristics that are used inputs for various models. Examples of typical characteristics are shown in Table 2.1.

Table 2.1: Typical physical and operational characteristics of level crossings used as inputs to SRPTs

Physical characteristics	Operational characteristics
• Visibility of the level crossing from the road approach • Gradient of the road approach • Width of the road at the level crossing • Proximity to other road intersections • Construction of road surface (paved, or unpaved)	• Maximum train speed over the level crossing • Proportion of freight train traversals • Longest approach warning time • Heavy vehicle proportion • Number of operational rail lines

RSSB note that there is a variety of methods used to calculate outputs from the various inputs used by the models. They classify the method of calculation in three categories. The first category they describe as *parameter gate* tools which use "*simple parameters as decision guides*". These tools do no produced a risk estimate, rather they provide guidance for the selection of warning devices that are required at each level crossing. The second category are described by RSSB as *weighting factor* tools which perform some calculation on the inputs to produce a numerical output which presents some indication of safety risk. These numerical values may be relative scores to provide a ranking of level crossings against each other, and thereby help create a priority listing for interventions. Conversely the numerical results may represent absolute risk predictions of the numbers of collisions that can be expected at a level crossing in the future. Finally there are the *statistically driven* tools use statistical methods from analysis of prior rates of collision to provide risk estimates of future rates of collision.

The report notes that the output of an SRPT does not necessarily provide the only data used to determine the warning devices required at a level crossing, nor for prioritising interventions, rather "*the model itself will provide only a part of the overall process for decision-making on level crossings*". Figure 2.5 has been adapted from RSSB to show the flow of data from the inputs to the model to the decision-making process and emphasises that the output from the SRPT is an input to the evaluation and decision-making that leads to interventions at level crossing.

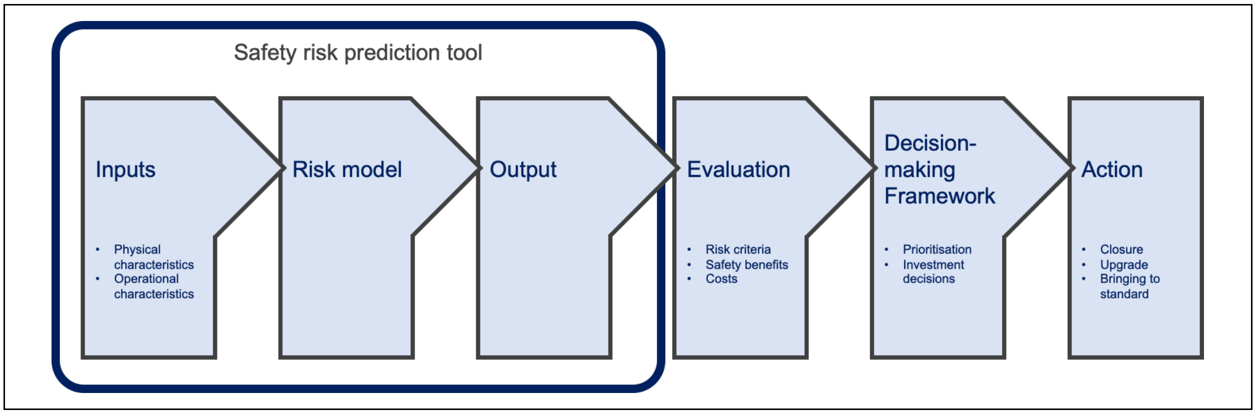


Figure 2.5: Application of SRPTs in decision-making process, adapted from RSSB (2007)

2.5.2 Use of traffic models in tools

Information about some of the tools was provided to RSSB through private agreements with the owners of the SRPTs which means that the details of the tools are not publicly available. One area where there is information regarding all of the SRPTs is that each uses some traffic model that describes how safety risk varies with varying road traffic and train volumes at a level crossing. The review identified that most of the SRPTs consider that safety risk varies in proportion to *traffic moment*: the product of the number of vehicular road users and trains in a given period (usually one day). Table 2.2 shows a summary of each of the tools studied by RSSB and the underlying traffic models for each; a dash (–) indicates that the traffic model is not specified. Notes are provided after the table.

Table 2.2: Summary of traffic models identified in RSSB (2007)

Country	Tool	Traffic model	Notes
GB (rail)	Automatic Level Crossings Model	–	(1)
	All Level Crossings Risk Model (ALCRM)	Stott's hypothesis	
	Event Window Model	–	
GB (highways)	COBA Junction Model	–	(2)
Australia	Risk Based Scoring System (RBSS)	Traffic moment	(3)
	Australian Level Crossing Assessment Model (ALCAM)	Traffic moment	
	RAAILc	–	
Canada	Collision Prediction Model	Traffic moment	
	GradeX	–	(4)
India	Train Vehicle Unit	Traffic moment	
Ireland	Network Risk Model	Traffic moment	
	Level Crossing Prioritisation Tool	Traffic moment	
Japan	Closed Road Traffic Indicator	Traffic moment	
	Level Crossing Danger Index	Sum of road traffic volume, train volume, and passengers per day	(5)
Northern Ireland	Risk Assessment and Investment Appraisal	–	
New Zealand	Product Assessment	Modified traffic moment	(6)
	Accident Prediction Model	–	(7)
Russia	Rail and Road Intensity Matrix	Non-linear, categorical risk prediction	(8)
Spain	Crossing categorising criteria	Traffic moment	
	FMEA method	Traffic moment	
Sweden	Factors to determine crossing protection	Traffic moment	
USA	APF and SPF	Traffic moment	
	GradeDEC.NET	Traffic moment	

Notes for Table 2.2:

(1) RSSB (2007) state: "*there is no detailed specification or mathematical description of the model*".

(2) RSSB (2007) notes that the traffic model is a non-linear function of road user volumes, but the exact form is not specified.

(3) The traffic model is not specified in RSSB (2007), however it is noted that the model does use traffic moment.

(4) The tool does not appear to normalise collision risk predictions by traffic volumes, rather the risk calculation appears to consider prior collision history at each level crossing.

(5) RSSB (2007) notes that the model: "*is a traffic moment, weighted by exposure to passengers as a measure of potential consequence severity, and also weighted by accident history*".

(6) The tool considers uses a traffic model that is similar to traffic moment, but the model is weighted for the time of day that rail traffic operates.

(7) Similar to Note (2), the model based on an unspecified non-linear function of road user volumes.

(8) The tool does not provide a risk prediction *per se* rather it provides risk categorisation based on road traffic and train volumes.

2.6 Proliferation of SRPTs

During the study of SRPTs, it is notable that some countries appear to have more than one tool for determining safety risk, the reason for this is unclear. It might be assumed that the different tools produce different results: otherwise there would be no need for more than one tool. Taken together with the lack of information on validation of any of the tools, it must be assumed that there is uncertainty regarding the most appropriate method for calculating safety risk. It can be expected that if a railway authority were to discover that their tool is highly accurate in risk prediction – and therefore a valuable method for allocating resources to public – it would be in the interests of the railway to make this information publicly available. Furthermore, if it were demonstrated that there were an accurate tool that was universally applicable then, over time, the tool would be adopted by all railways from all countries. The proliferation of different tools is suggestive of a situation where no tools has been developed that has generally good predictive accuracy.

2.7 GB Railways' ALCRM

Whilst most of the SRPTs use traffic moment as the traffic model, it is notable that the tool used by the GB, the All Level Crossing Risk Model (ALCRM) uses a different model. As noted in RSSB (2007) the ALCRM does not assume "*that risk is proportional to traffic moment (as with all other models)*", rather the tool uses a hypothesis proposed by Stott in 1987 to determine how safety risk varies with changing road traffic volumes. In reviewing the efficiency of the ALCRM in 2007, Baker and Heavisides (2007) state that abandoning a simple proportional model and adopting Stott's hypothesis "*has caused a significant re-appraisal of which are the highest-risk level crossings in GB. Some crossings are now shown by the ALCRM to be relatively higher risk than previously thought, while other more busy crossings may actually be safer*". Despite the fact that the use of Stott's hypothesis has led to a re-evaluation of safety risk, no information is provided to show whether incorporation of the hypothesis leads to a better match between predicted and observed collision rates.

Stott's hypothesis regarding the effect that varying traffic volumes have on safety risk is described below. However, as for a number of the SRPTs, the choice of traffic model is the only information that is available on the method of calculation used in the ALCRM.

Stott's hypothesis describes a non-linear relationship between road traffic volume and safety risk, however it is notable that the hypothesis does not describe any relationship between train volume and safety risk. Rather it appears that the hypothesis considers that the risk is constant *per train* and therefore that safety risk will increase proportionally with train volume in the same way as it does with the traffic moment model. It is not clear that this such an assumption is necessarily valid, the description of the Australian SRPT (the ALCAM) provided by Hughes (2002) shows that the tool considers that train volume can contribute to safety risk at level crossings: at low train volume the ALCAM considers that road users who regularly use the level crossing may become used to the idea that trains are rare and will not expect a train arrival and consequently will fail to adequately prepare on the approach to the level crossing. Conversely where there is a high volume of trains at a level crossing, regular road users may tire of waiting for trains and may feel encouraged to cross in front of an approaching train. If the effects are real, it is not clear the degree to which each affects safety risk; it is theoretically possible that the effects perfectly balance each other in a way that makes safety risk exactly proportional to train volume. However there is not information available on whether such an assumption is reasonable.

2.8 The United States Department of Transportation SRPT

The RSSB review is the most comprehensive source found for reviewing the various SRPTs that are in use, however it is not complete. Wullems *et al.* (2013) describe the SRPT used by the United States Department of Transport which is underpinned by the traffic model developed by Peabody and Dimmick in 1941. The original source of this model can no longer be found, although it is described in the Railroad-Highway Grade Crossing Handbook (United States Department of Transportation, 2007). The Peabody Dimmick model was developed based on a study of collisions that occurred in 29 states over a five-year period. The Peabody Dimmick model describes a non-linear relationship between road traffic volume and safety risk, although the relationship is different to that described by Stott's model. Again, the model neglects to mention any relationship between train volume and safety risk and it again appears that the tool assumes that safety risk is entirely proportional to train volume.

2.9 Summary of traffic models used in SRPTs

The review of traffic models identified six models used in the various SPRT, these are summarised in Table 2.3. Amongst these models, detailed information exists for three of the models to be studied further, being:

- traffic moment;
- Stott's hypothesis; and
- Peabody Dimmick's model.

Table 2.3: Traffic models used in SPRTs

Traffic model	Application	Notes
Traffic moment	Used in 12 SRPTs	–
Sum of road traffic volume, train volume, and passengers per day	Used in Japanese *Level Crossing Danger Index*	Requires knowledge of the number of passengers on a train
Stott's hypothesis	Used in GB (rail) *All Level Crossings Risk Model* (ALCRM)	–
Non-linear, categorical risk prediction	Used in Russian *Rail and Road Intensity Matrix*	–
Modified traffic moment	Used in New Zealand *Product Assessment*	Requires knowledge of the proportion of trains operating at night
Peabody Dimmick	Used in United States' *DoT model*	–

2.10 Classification of proximate causes of collisions in the ALCAM

An exception to the general dearth of information regarding mechanisms of calculation is the Australian Level Crossing Assessment Model (ALCAM) described by Hughes (2002). A notable feature of the ALCAM which, from the available information, appears to be unique, is the categorisation of proximate causes of collisions at level crossings, which are dubbed *accident mechanisms*. Examples of accident mechanism used in the ALCAM include:

- road users being distracted by adjacent distractions which leads to a collision;
- road users being warned too late of the need to stop and consequently not having sufficient time to avoid a collision; and
- road users queuing over a level crossing.

The ALCAM considers 19 accident mechanisms, which are grouped into three categories. The categories describe cases where road users are:

- unaware of a level crossing, or are aware of the level crossing but unaware of the need to stop;
- aware of a level crossing but are unable to take action to avoid a collision; and
- aware of the level crossing and able to avoid a collision but misjudge their traverse in a way that results in a collision. In this case the misjudgement includes cases where road users wilfully breach active warning devices.

This categorisation is intuitively appealing in that it accords with a simple classification of allocation of blame, however it does not appear to be supported by either theory nor observation. In this regard the classification may be viewed as a starting point for further research, however in the time since publication of Hughes's work, no such studies appear to have been undertaken.

2.11 Traffic models

This section describes in detail the three traffic models that are studied further in this research.

2.11.1 Traffic moment

Baker and Heavisides (2007) define traffic moment as "*the product of the number of trains and number of level crossing users*". The same measure is dubbed the *VT product* by Hughes (2002) who explains it as "*the product of the daily number of road vehicle crossings (V) and the daily number of train crossings (T)*" and goes on to assert: "*the expected number of accidents at a level crossing is proportional to the probability* *that a single road user on a single crossing will be involved in an accident multiplied by the number of chances that there are for accidents to occur*". It would appear that Hughes considers it axiomatic that traffic moment is a natural normaliser for level crossing collisions and may be the reason that 12 SRPTs shown in Table 2.2 use traffic moment was also used as the normaliser for collisions. Whilst it appears that there is support for the concept of using traffic moment as a normaliser, there is no evidence for why this should be the case. Traffic moment is therefore a *predictive* model is that is was derived by means of logical reasoning, rather than collection of empirical data collection.

2.11.2 Stott's hypothesis

An alternative predictive traffic model was proposed by Stott (1987). Stott states "*at first sight it might seem intuitive that increasing the [road] traffic volume would bring proportionately more collisions*" but goes on to propose a more complex hypothesis: "*assume, first, that on the great majority of occasions, drivers stop when the lights show...second, assume that if a vehicle does stop it acts effectively as a barrier to all following vehicles*". The hypothesis further proposes that if a road user breaches the holding point at a level crossing soon after the active warning has started, the road user has the opportunity to recover before any trains arrive, and is therefore unlikely to be involved in a collision. It is only road users who breach the holding point just before the arrival of a train who are likely to be involved in collisions. Furthermore, the higher the level of road use at a level crossing, the more likely it is that road users will arrive at the level crossing early and stop correctly, forming a barrier to those who arrive later. In summary, Stott hypothesised that: for zero road users there is a zero likelihood of collision; the likelihood of a collision per activation of the level crossing warning initially increases as the number of road users increases; however for high numbers of road users the likelihood of a collision decreases to a low value. Stated mathematically, the hypothesis is that the number of collisions at a level crossing varies in accordance with a distribution over the number of road users at the level crossing in the form:

$$\text{probability of a collision per activation of the level crossing warning} = P_C \times e^{-m(T-t)} - e^{-mT}$$

Where:

P_C	is the probability that a road user arriving at the holding point of a level crossing fails to stop in accordance with the warning;
e	is the base of the natural logarithm;
m	is the rate that road users arrive at the level crossing;
T	is the time between the active warning starting and a train arriving at the level crossing; and
t	is the minimum time required for a road user to recover from a breach.

Stott proposed the following values for calculating the hypothesised collision rate: $P_C = 0.0001$; $T = 40$ seconds; $t = 2$ seconds. Figure 2.6 provides a graphical representation of the hypothesised distribution presented by Stott.

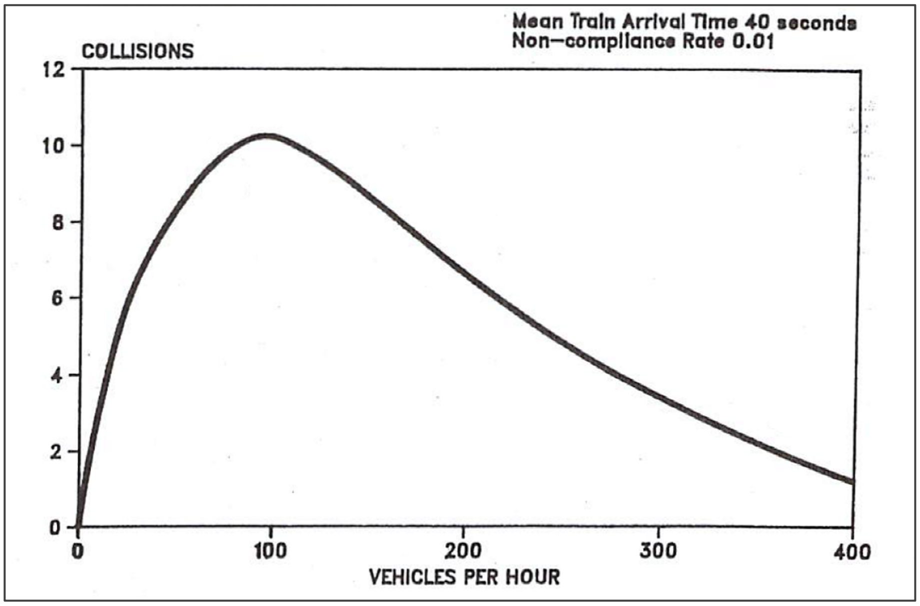


Figure 2.6: Schematic representation of hypothesised collision rate over road traffic volume, adapted from Stott (1987)

In his hypothesis, Stott presumed that road traffic arrivals would not be constant during a 24-hour period and, although it is not explicitly stated in his work, it is implied that to determine an overall rate of collisions it is necessary to calculate the expected collision rate in each hour and sum the results for a full day. The traffic distribution presumed by Stott is shown in Table 2 of Annex C of his paper (1987) and reproduced in Table 2.4 below.

Table 2.4: Distribution of road user arrivals throughout a day copied from Stott (1987)

Time of day	Proportion of road user arrivals at a level crossing
00:00 hours to 07:00 hours	0%
07:00 hours to 08:00 hours	5%
08:00 hours to 09:00 hours	7%
09:00 hours to 10:00 hours	6%
10:00 hours to 11:00 hours	7%
11:00 hours to 12:00 hours	7%
12:00 hours to 13:00 hours	6%
13:00 hours to 14:00 hours	6%
14:00 hours to 15:00 hours	7%
15:00 hours to 16:00 hours	8%
16:00 hours to 17:00 hours	9%
17:00 hours to 18:00 hours	8%
18:00 hours to 19:00 hours	5%
19:00 hours to 20:00 hours	5%
20:00 hours to 21:00 hours	5%
21:00 hours to 22:00 hours	4%
22:00 hours to 23:00 hours	3%
23:00 hours to 00:00 hours	2%

It is notable that while Stott considers varying road traffic volume, it is considered that train arrivals are constant throughout the day.

2.11.3 Peabody Dimmick model

Faghri and Demetsky (1986) state that the Peabody Dimmick model was first published in 1941 "*based on 5 years of accident data from rural crossings in 29 states of the USA*". Unlike the other two models, the Peabody Dimmick model is therefore *descriptive* in that it is based on empirical observation of collisions. The model gives the expected number of collisions at a level crossing in a five-year period as being:

$$A5 = Iu + K \qquad (1)$$

In this definition:

K is a parameter obtained from the a graph provided in the text; and

$$Iu = \frac{1.28(V^{0.170} \times T^{0.151})}{P^{0.171}} \qquad (2)$$

Where:

V is the annual average daily road traffic;

T is the annual average rail traffic; and

P is the protection coefficient which varies for the class of level crossing.

Values for P are provided by the US Department of Transportation (2007); the values applicable to this study are shown in Table 2.5.

Table 2.5: Values for *P* provide by the US Department of Transportation (2007) used in this study

Type of warning devices	Application to this study	Value for *P*
Signs	Passive warning devices	1.65
Automatic gates	Automatic warning devices	2.56
Watchman, 24 hours	Railway-controlled warning	2.52

The exact calculation for K is not defined parametrically in the original source. Figure 2.7 is reproduced from Wullems *et al.* (2013) and graphically shows the relationship between K and Iu.

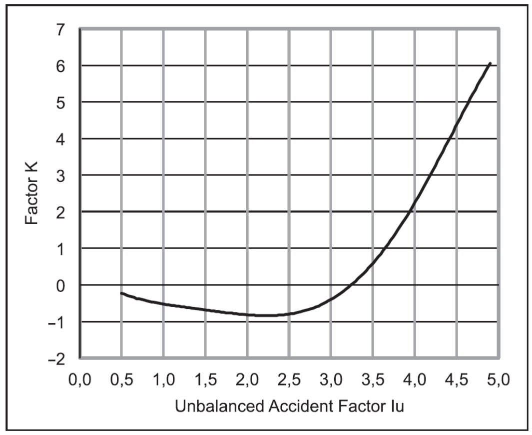


Figure 2.7: Peabody Dimmick relationship between *K* and *Iu* reproduced from Wullems *et al.* (2013)

In order to allow calculation from the formula, Wullems *et al.* performed a polynomial regression to determine:

$$K = -0.0329\, Iu^5 + 0.3996\, Iu^4 - 1.604\, Iu^3 + 2.9503\, Iu^2 - 1.891\, Iu + 0.654 \qquad (3)$$

Inserting the result from Equation (3) into Equation (1) gives the full calculation for predicted number of collisions. This form of the equation was used by Wullems *et al.* (2013) in their work and is the form that will be used for further study in this work.

2.12 Applicability of traffic models for different classes of level crossing

Traffic moment is the most commonly used normaliser in SRPTs. In describing the Australian SRPT (ALCAM, 2007), Hughes (2002) explains that the method of calculation assumes there is a constant probability that any road user will fail to take correct action on the approach to a particular level crossing. The probability of a road user error may vary between level crossings and can depend on features such as the presence and type of warning device at each level crossing. Assuming that there is a fixed probability of each road user committing an error, the probability of a collision occurring depends on whether there is a train on, or approaching, the level crossing at the time the error is made. Therefore the probability of a collision increases linearly with the number of train arrivals and the number of road users. Applying this reasoning, the traffic moment model therefore applies to all level crossings.

Stott's traffic model was proposed as part of a review of automatic open level crossings (Stott, 1987), which are level crossings with active warning devices but no barriers nor boom gates over the road approach. As such it is possible that the model was never intended to be applied to other classes of level crossing. In proposing the model, discusses the behaviour of vehicular road users approaching a level crossing. Stott presumes that the first road user to stop at the holding point of a level crossing will cause an obstruction on the road that will cause all following road users to stop. He argues that a collision can occur only if the first road user fails to stop correctly: failures of subsequent road users may result in collisions with other road users but will not result in a collision with a train. In making this argument it is not relevant why the first road user stopped: whether it was because of the road user looked for an approaching train, or because of the warning provided by active devices. Furthermore Baker and Heavisides (2007) describe the adoption of Stott's traffic model for vehicular level crossings in the British SRPT, seemingly for all classes of level crossing. Whilst Stott proposed his model for only a particular class of level crossing, it seems that it can be, and is, used on other classes of vehicular level crossing.

The model proposed by Peabody and Dimmick was based on the number of collisions observed at level crossings in Illinois during the 1930s and includes a factor (*P*) for the class of level crossing. Different values of *P* are provided for level crossings which have stop signs (passive warning devices), those that have automatic gates, and those that have a watchman. Therefore it is possible to apply the model for different classes of level crossing. For railway-controlled level crossings in Britain the role of a signaller in manually operating a level crossing's warning system goes beyond that of being only a watchman, since the signaller also clears the signal to allow the train to approach the level crossing (Evans and Hughes, 2019), as such it is possible that application of the model for this class of level crossing is beyond the scope intended by Peabody and Dimmick when they created their model.

2.13 Point estimates predictions of traffic models

It is noted that the traffic models provide point estimates of the expected numbers of collisions that will occur at a particular level crossing. As such, it must be expected that all traffic models will be incorrect in some way: the rates of collision predicted by the models will always differ from the observed rate of collisions due to natural randomness in the environment and the factors that cause collisions.

Rather than the models giving only point estimates, an alternative approach would be for a model to probabilistic predictions for the rate of collisions that can be expected at a level crossing. Such probabilistic models could define coefficients of variation for the predicted numbers of collisions: in this way the models could be considered to be mathematically more correct in that they would not produce results that can be shown to be incorrect. Whilst such an approach is appealing mathematically, it is not clear that it would necessarily provide any advantage for safety management.

The various traffic models, and the SRPTs that use them, have been created to provide road and rail authorities with a reliable method to determine which, if any, level crossings should be prioritised for investment to reduce safety risk. RSSB (2007) note that some SRPTs produce absolute predictions of the numbers of collisions expected at each level crossing, whereas other produce relative rankings of level crossings against each other. In either case, the tools can be used to prioritise which level crossings require more immediate intervention to reduce risk.

Railway safety legislation in the United Kingdom requires that safety risks are reduced to a level that is as low as is reasonably practicable. At the current time, the railways do not appear to be claiming that the overall safety risk from all level crossings has been reduced so far that no further interventions are required. Therefore a traffic model or SRPT will be useful to the railway if it is able to provide an accurate prediction of which level crossings have present the highest safety risk. It is likely that a probabilistic traffic model will be more complex to create that one that provides only point estimates; therefore it is understandable that the models adopted by road and railway authorities use point estimates.

2.14 Data-driven railway safety management

RSSB's research from 2007 remains one of the most recent works, as well as the prominent work collating information on the various SRPTs in use around the world. When considering the traffic models that underpin the tools, it is notable that no source can be found for the derivation of traffic moment as a normaliser for traffic volume. When considering the other main traffic models Stott's hypothesis was published in 1987 and the Peabody Dimmick model was developed in the 1930s. Recently there has been a dearth of development of SRPTs and their traffic models.

This stagnation of SRPTs is at odds with the trend in advancement of information technology in general and specifically data-driven safety management. An emerging research theme on *big data risk analysis* has focussed on collecting voluminous amounts of data on railway operations in order to understand not only accidents that occur on the railway (Rashidy *et al.*, 2018; van Gulijk *et al.*, 2016) but also to understand the hazards on the railway so that the underlying causes can be removed before an accident occurs (Hughes *et al.*, 2016; van Gulijk and McCullogh, 2019). The research concentrates on creating descriptive models of hazards and accident causation. With modern information technology systems the models can become significantly more complicated than the SRPTs used for level crossing safety as they describe subtle effects in the underlying mechanisms of accident causation (Hanea *et al.*, 2012). As well as collecting data on accidents and hazards, the work is also focussed on collecting data on successful completion of safety critical tasks that do not lead to accidents (Hollnagel, 2018; Rashidy *et al.*, 2018). In principle the data could be collected and processed to update the models in near real-time.

It is clear that there is an opportunity to apply the findings of this research to science of level crossing safety risk management where substantive research does not appear to have occurred since 2007 and in one case an underlying traffic model that was developed in the 1930s is still being used for modern safety risk management.

2.15 Machine learning

The founding assumption for use of any SRPT is that there is a relationship between the physical and operational characteristics and the safety risk – and therefore the rate of collisions – at level crossings. Traditional techniques for determining the degree of correlation use statistical methods such as logistic or logit regression. However such techniques are limited in their application: Molnar (2018) notes "*linear regression and logistic regression models fail in situations where the relationship between features and outcome is nonlinear or where features interact with each other*". There is no reason to assume that the physical and operational characteristics that affect safety risk at level crossings should be either linear, or independent of each other. For example the safety risk may not be affected simply by the number of heavy goods vehicles (HGVs) traversing a level crossing, instead the risk may increase when there is a combination of HGVs and a short warning time. Furthermore it is possible that the hazards do not correlate linearly with safety risk: hypothetically it is possible that when there is a short warning time, road users respond by taking extra care before traversing; where there is a long warning time there is a low safety risk. Perhaps the greatest safety risk occurs at only some intermediate value of warning time. An exhaustive analysis of all combinations of all possible non-linear correlations would be intractable using statistical methods, however the emerging science of machine learning is well suited to the task of detecting correlations between variables. At present there is no consensus in the literature on a definition of the term machine learning, although there is a broad agreement.

Definitions from amongst the literature include:

- "*a set of methods that can automatically detect patterns in data, and then use the uncovered patterns to predict future data, or to perform other kinds of decision making under uncertainty*" Murphy (2012),
- "*If a system can improve its performance by performing a certain process, it is learning*" Li, Zhang and Zhang (2017),
- "*learning can change the performance of the system*" (ibid),
- "*learning is the ability to change according to external stimuli and remembering most of all previous experience*" Bonaccorso (2017).

These definitions do not agree overall, for example it appears that Li *et al*. believe learning occurs only when a process is being repeatedly performed, whereas Murphy implies that learning can occur during a one-off analysis of data. Regardless of these differences, the literature all describe machine learning as a process where software computes an output using a method that was not previously defined by a programmer; instead the calculation method depends on observation of sample data. As Bonaccorso implies, machine learning can be a valuable tool to produce meaningful outputs when the method of calculation cannot readily be determined by a human programmer.

Compared with statistical techniques, machine learning methods can be computationally expensive (Bzdok *et al.*, 2018), however the increasing power of computer systems mean that it is now practicable to perform complex computational tasks on personal computers which previously would have been too taken too many computing resources to be practicable (Golio, 2015). Consequently there has been a recent surge in the application of machine learning methods, a review of internet searches using data provided by Google Trends online data profiling service (2019) shows that the number of search requests worldwide has more than quadrupled since 2013. With modern computers continuing to become more powerful, the range of problems that can be tackled by software is increasing; tasks that were infeasible with earlier software techniques are possible with modern computers performing machine learning. Bonaccorso (2017) notes "*without machine learning there are still many tasks that seem far out of computer domain*" and goes on to list examples such as: spam filtering, natural language processing, and visual tracking with a webcam. One aspect of machine learning where there is very nearly a consensus is that there are three main types of machine learning methods: *supervised learning*, *unsupervised learning*, and *reinforcement learning*. Again, the exact definitions vary between the sources, but there is common agreement on the meaning of the terms.

Supervised machine learning uses software to create a method of calculation from observation of data with a number of properties that are classified into one or more classes. The role of supervised machine learning algorithm is to identify patterns in the properties of an object (within the nomenclature of machine learning, these are often referred to as *features*) that allow reliable prediction of the classification (known as *labels*). Alpaydin (2009) provides an example of classifying whether or not a vehicle is a *family car* based on the vehicle's price and engine power. In the example it is presumed that there is a certain range of prices and a certain range of engine powers that are consistent with a vehicle being a family car, outside of either range the vehicle cannot be classified as such (perhaps it would be better classified as a truck or a motorcycle). Supervised machine learning determines a calculation method from data which is already classified: sometimes called *labelled* data (Stamp, 2017). These input data are split into two sets: the *training set* and the *testing set*. The first part of the machine learning process attempts to identify patterns in the training set that allow classification of the data based on the properties. Once patterns have been identified, the accuracy of the classification is tested using the testing set. Since the data in the testing set are already labelled, it is possible to compare the classification determined from the machine learning algorithm with the prior classification. Once an acceptable level of accuracy has been obtained using the labelled data, the same algorithm is then applied to unlabelled data to determine classifications. The process assumes that the unlabelled data are sufficiently similar to the labelled data that the accuracy of classification prediction will be approximately the same between the two cases.

Unsupervised learning is applied where the data are not labelled; Stamp (2017) describes the process as "*most useful in cases where we don't know much about the data*" as "*it can help us determine possible structure* (in the data)". Kyan *et al.* (2014) provide an example of being able to differentiate animals into categories – insects or birds – based on the animals' social dynamics: swarming, hive-making, or flocking. It is important to note that unsupervised machine learning algorithms do not label the data; rather they identify separations in the data – for example there is a clear separation between insects and birds when the properties of social behaviour are considered. It remains the task of humans infer meaning from the classifications and provide labels for the groups that have been created. Having created a method of separating data, new data can be classified into one of the existing classes: a new animal can be classified as either a bird or an insect using the same algorithm.

Reinforcement learning is described by Alpaydin (2009) as:

> *...the learner is a decision-making agent that takes actions in an environment and receives reward (or penalty) for its actions in trying to solve a problem. After a set of trial-and-error runs, it should learn the best policy.*

Invariably the literature exemplify reinforcement learning with cases of software undergoing a learning process to achieve improved scores when playing simple video games or board games such as noughts-and-crosses or chess (Amato and Shani, 2010). During the learning process, a large number of games are played with the software initially playing moves at random. If a game ends with the software winning then the moves played during the winning game are preferred during future games. Whilst playing many games, the software continues to use a stochastic process to select which move to play, where the experience from previous games is considered and moves that resulted in previous wins are more likely to be selected than those that previously resulted in losses. Reinforcement learning requires a large amount of repetition and it is intended that the output of the software gradually improves over time to obtained a desired result.

In summarising these three types of machine learning, a diagram that has become widespread in the literature is shown in Figure 2.8; many sources have similar diagrams, this version appears in Educba (*n.d.*) on their website (https://www.educba.com/machine-learning-algorithms/, retrieved on 21 November 2019).

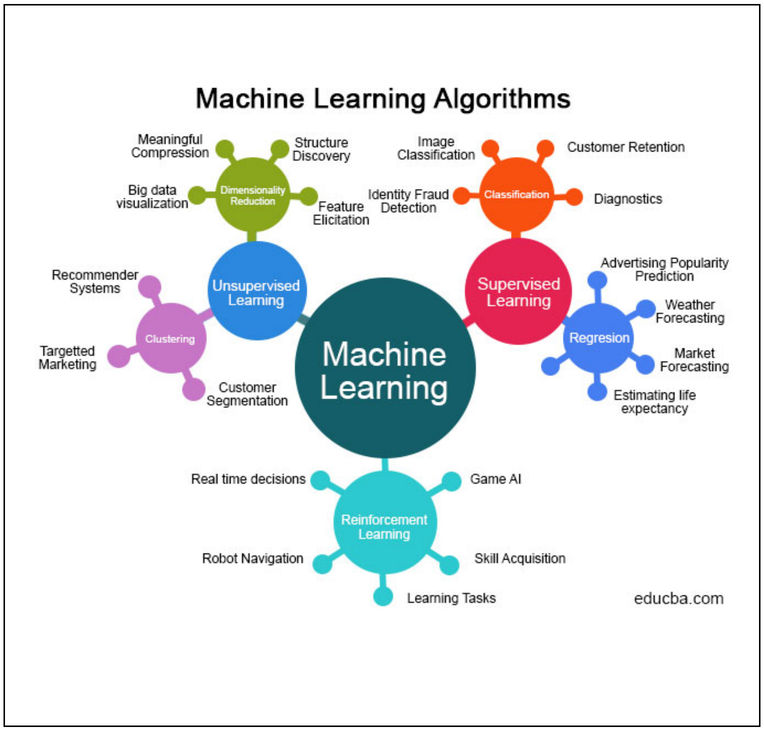


Figure 2.8: Overview of machine learning categories, reproduced from Educba (*n.d.*)

An important aspect of machine learning classification algorithms is that, given an input, an output will always be created, however the output may not always be considered to be *correct* when compared to the classification that would be provided by a human. Provost and Fawcett (2013) explain:

> *…two related fundamental concepts of data science: generalization and overfitting. Generalization is the property of a model or modeling process, whereby the model applies to data that were not used to build the model… Overfitting is the tendency of data mining procedures to tailor models to the training data, at the expense of generalization to previously unseen data points.*

For example, based on the social behaviours of an animal, a machine learning algorithm may assign a particular animal to be in the same category as insects, when the animal is, in fact, a bird. Inaccuracy of output is an inherent feature of machine learning algorithms. In many reinforcement learning applications, the first output provided by the algorithm is entirely random. It would therefore be unlikely that the output is considered correct, rather it is expected that the accuracy of the algorithm will improve over repetition until an acceptable level of accuracy is obtained.

Within the scheme presented in Figure 2.8, the task of identifying how characteristics of level crossings correspond with safety risk is a *classification* problem to determine which characteristics correlate with observed rates of collision. Amongst the available techniques, some are more accurate at linear classification, while others are more suited to nonlinear classification. Again, the literature are not unanimous on the machine learning methods that should be applied for classification, although there is broad agreement. A number of comprehensive sources were reviewed to identify candidate methods. Table 2.6 shows results from a sample of authors that typify the broad range of literature available on machine learning; in the table a tick mark (✓) shows where authors have proposed a method for machine learning classification tasks.

Table 2.6: Machine learning methods for classification as identified by various sources of typical sources

	Authors			
Machine learning method	**Aggarwal (2014)**	**Wickham (2018)**	**Lantz (2013)**	**Awad and Khanna (2015)**
Decision trees	✓	–	✓	–
Random forest	–	✓	–	–
Naive Bayes	–	✓	✓	–
Artificial neural networks	✓	–	✓	✓
Support vector machines	✓	✓	✓	✓
k-nearest neighbours	–	✓	✓	–

Each of the candidate methods identified in the literature review are described in the sections below.

2.15.1 Decision trees

The *decision tree* method creates a flowchart where each fork in the chart represents a classification criterion based on the input data. Repeated forking leads to a single classification for each data item. Figure 2.9 shows an example decision tree that was created from analysis of data to determine tax payments based on properties of individuals, reproduced from Ayyadevara (2018).

	categorical	categorical	continuous	class
Tid	Refund	Marital Status	Taxable Income	Cheat
1	Yes	Single	125K	No
2	No	Married	100K	No
3	No	Single	70K	No
4	Yes	Married	120K	No
5	No	Divorced	95K	Yes
6	No	Married	60K	No
7	Yes	Divorced	220K	No
8	No	Single	85K	Yes
9	No	Married	75K	No
10	No	Single	90K	Yes

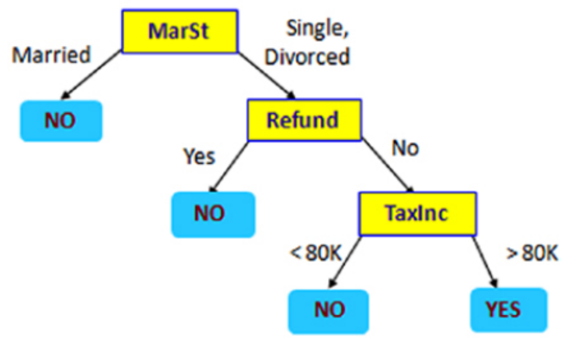


Figure 2.9: Example decision tree derived from data, reproduced from Ayyadevara (2018)

To create a decision tree, it is necessary to determine which properties of the data form the decision forks and in which order they should be applied. An algorithm to create a decision tree from data – the *Iterative Dichotemiser 3* (ID3) method – was described by Quinlan (1986). The ID3 algorithm calculates the entropy of every attribute present in a given dataset and splits the data so that the resulting sets have the lowest entropy. In considering this approach Mohammed *et al.* (2016) note that this method is susceptible to over-partitioning, *i.e.* splitting data into more classes than necessary, due to the assumption that entropy reduction is the best measure to determine the best split.

Whilst a decision tree created using the ID3 would be optimal in terms of entropy reduction at each fork, the process to create such a tree can be computationally expensive where there are a large number of properties on the data. An alternative to the ID3 algorithm is the *C4.5* algorithm which uses the *information ratio* metric instead of entropy to derive forks in the dataset (Fran and Hall, 2011; Larose and Larose, 2014). The C4.5 algorithm normalises the entropy reduction from any fork using the number of datasets which would result from that fork. This approach removes the problem of the ID3 algorithm where a property which has a large amount of variability but is not necessarily useful for making real-world classifications is given weight due to the high reduction of entropy in the resulting sets. For example, a set of patient data in a hospital may contain properties such as patient names, symptoms, and diagnoses. By selecting a fork for greatest entropy reduction, the ID3 algorithm might select patient name as the most meaningful attribute for prediction of a rare diagnosis, as it is the attribute which results in the highest entropy reduction when all names are placed in separate categories.

For application in this study, a decision tree could be constructed to identify the features of level crossings that correspond with observed collisions. It would be possible to use training data that describe the features of level crossings, together with a label indicating either the number of collisions that have been recorded at the level crossing, or the normalised rate of level crossing collisions. Since decision trees predict membership of categories, using the method in this way would require level crossings to be categorised in terms of their history of collisions. In the simplest case, level crossings could be categorised by whether there is any history of collisions or not. A more detailed approach would be to categorise the numbers of collisions: for example level crossings that have a collision of one or two collisions, and those that have a history of more than two collisions.

Training data describing the features of each level crossing would be applied to the method. In theory the features of a level crossing could be any numeric or categorical value such as the maximum speed of trains traversing the level crossing, or the categories of trains, such as passenger trains or freight trains. Details regarding other features of the level crossing – such as the presence of nearby distractions for road users – could be included if these data could be coded as either numeric or categorical values. For example Boolean values could be used to indicate whether there were nearby road junctions, or signage. The decision tree method could be applied to construct a tree that describes the features of level crossings that correlate with categories of observed collisions and those that correlate with no history of collisions.

Data on features of a level crossing could be applied to the resulting decision tree to determine the category of collision, or whether no collisions are predicted at an individual level crossing. Furthermore the decision tree can be examined to understand the method used to predict rates of collision. Features nearer the root of the tree are usually the ones that have the greatest impact on the prediction, and therefore are the ones that have the greatest impact on the collision prediction. In cases where the resultant decision tree produces a high degree of accuracy – even if it is not completely accurate – examination of the resultant tree may provide useful information on the features of level crossings that correlate with a history of collisions. The resulting tree could therefore be used to guide safety interventions at level crossings, even in cases where the trees predictions are not completely accurate.

2.15.2 Random forests

The *random forest* method was developed to address the problem of overfitting by decisions trees by creating multiple decision trees where each tree uses only a subset of the properties of the data (Berk, 2008). By using only a subset of the properties it is not possible for the algorithm to overfit the data.

Having created a number of decision trees (a *forest*) the trees are each tested for the predictive accuracy. A common method using random forests is to select a number of the trees that overall provide the most accurate predictions on the training data. Unseen data are then applied to the subset of decision trees and are classified based on the majority outcome from all the trees used (Grus, 2019).

The random forest method is based on the decision tree method and can be used in any application where it is possible to create a decision tree. Since the random forest method creates a large number of decision trees, the trees in the forest consider different features of level crossings and have different criteria for selecting the categorisation of an individual level crossing. This approach may produce more accurate predictions than the decision tree method, however examination of the many trees created using this method will be more difficult than examination of the single tree created by the decision tree method. By considering only a subset of the features of level crossings, the various trees in the random forest will consider the features in different ways. A feature that is near to the root in one tree may not be near the root in another tree: in other trees the feature does not occur at all. The larger the number of trees the more difficult it will be to discern the main features that affect rates of collision. As such, whilst the random forest method might produce more accurate results than the decision tree method, it may be less useful for determine appropriate safety interventions.

2.15.3 Naive Bayes

The *naive Bayes* method uses Bayes' probability theorem to compute the posterior probability that a data item belongs to a particular class based on its features (Aggarwal, 2014). The probability of an item belonging to each class is calculated and the method assigns the item to the class with the largest probability. An advantage of this method is that an explicit probability estimate is provided for each class, therefore it is possible to determine the confidence with which a classification has been made. Another advantage of the method is that is scales well to large sets of data and can be used in cases where there are many features in the data (Witten and Frank, 2002).

The disadvantage of the naive Bayes method that is most often cited in the literature is that the method assumes that the features of the data are independent are do not combine to affect the classification of an item (Gan, 2020). As noted above it is unlikely that the features that affect level crossing safety risk are independent of each other and therefore application of the method for level crossing data could be expected to produce unreliable results. Nevertheless, experience of application on real-world data "*often results in classifiers that work well*" (Murphy, 2012), and therefore there is no reason not to test the method on level crossing data. A more profound problem with this method is the *zero-frequency problem* which occurs when a combination of features occurs in the testing set that has not been previously observed in the training set (Binmore, 2008). In these cases the naive Bayes method will produce a zero probability estimate for every class. Given the large number of features that can be used to describe a level crossing, and the wide variety of level crossings that occur, it is certain that there will be level crossings in the test set that contain combinations of features that have not previously been observed in the training data. For this reason, the naive Bayes method is not considered suitable for use in classifying level crossing data.

2.15.4 Artificial neural networks

Artificial neural networks (ANNs) are commonly described as "*the poster child of machine learning*" and are often considered to be synonymous with machine learning (Wujek, 2017; Whitehorn, 2017). Initially proposed in the 1950s, ANNs are an attempt to create logic in software that emulates the structure of the human brain (Haykin, 1994).

An ANN is made up of a number of *artificial neurons*, each neuron is capable of performing only limited computation, typically a single neuron will provide an output that is the sum of the weighted inputs compared to a threshold value. The neurons are arranged in layers so that the output from a neuron in one layer provides input to a number of neurons in the next layer. A large number of artificial neurons arranged in a network are able to provide complex output and are typically used for classification tasks.

An ANN has a large number of parameters defining the weights for each input to each neuron, initially the weights are set to random values. To train the network, a number of example input data are provided and the random network is used to provide a classification. Using the training data, the outputs from the network are compared with the expected outputs. Over repeated application of example inputs and comparison of the output with the expected results, the weights in the network are updated in an attempt to improve the general classification accuracy of the network. Two methods are used to update the weights: *backpropagation* and *genetic algorithms*. There is no consensus regarding which method, if either, should be preferred over the other (Garson, 1998; Guo and Uhrig, 1992; Fausett, 1994). As a relatively old technology amongst machine learning techniques, ANNs have been applied for many applications and the literature abound with descriptions of successful applications, especially for computer vision systems. Nevertheless there remain many cases where ANNs have not been found to be successful (Srivastava *et al.*, 2014).

For understanding safety risk at level crossings, it is possible to apply the ANN method in the same manner as decision trees would be applied (refer to Section 2.15.1). Data on features of level crossings could be provided as inputs either as numerical or categorical values. Since the output of an ANN is a value that corresponds to membership of a class, it would again be necessary to create classes to describe observed collisions at a level crossing. Again, this categorisation could be a simple Boolean value: whether a level crossing has a history of collisions or not; or more a complex categorisation that groups the number, or the rate, of observed collisions.

Unlike the decision tree method, ANNs are often described as being a *black box* method, indicating that the method of calculation cannot readily be interpreted. The method of calculation used by an ANN is coded in a large set of weights and threshold values in the network. A large number of values are calculated in parallel in the network; the results of many calculations are brought together to produce the final output. The parallel calculations within the network are determined by a process that simply aims to produce an accurate final result and, in many cases, will not be meaningful in terms of the real-world operations at level crossings. It is possible that the various parallel calculations even have contradictory effects on the output: one calculation may use a set of features to indicate a large number of collisions; whereas another calculation may use the same features to indicate no collisions. It is only when these results are combined in the final layers of the network that accurate results are achieved. As with random forests (refer to Section 2.15.2) it is possible that an ANN could produce results that have a high degree of accuracy, but the results are not more generally applicable to determining useful safety interventions at level crossing.

2.15.5 Support vector machines

Support vector machines are derived from the techniques used in linear regression. The simple example of a support vector machine would be to consider marks awarded to students who have taken an exam presented on a number line. Above a certain mark (say, 80%) the student is awarded a pass grade, below that mark, the student has failed. If there are two properties of the data to be considered then the decision boundary is not a single point but a line, for more properties the boundary becomes a plane or a hyperplane (Campbell and Ying, 2011). In general there can be a large number of boundaries that divide two datasets, the algorithm for development of a support vector machine calculates an optimal boundary that maximises the distance between the boundary and the datasets. Understanding support vector machines is relatively intuitive when there is a single boundary in the data, however real-life cases may not be so straight-forward. For example, when considering whether a person is a healthy weight, there are a range of values that can be considered healthy: values either above or below that range may be unhealthy. As such there is no single boundary upon which to make a decision. Support vector machines overcome this problem by transforming to the data to inflate the number of dimensions. in general, where there are sufficient dimensions in the data, it is always possible to identify a hyperplane that will divide the data in accordance with some classification (Lam *et al.*, 2012).

For assessing the features of level crossings that correspond with observed collisions, it is possible that support vector machines can be applied in the same manner as decision trees and ANNs (refer to Sections 2.15.1 and 2.15.4). Again it would be necessary to categorise the observed collisions and the method would create predictions within those categories. Like ANNs, support vector machines are black box methods: in principle it is possible for the software applying the method to show the method of prediction, however the results are not necessarily meaningful in terms of the physical features of a level crossing, and would be intractable to a human reader.

2.15.6 k-nearest neighbours

Based on labelled data, the *k-nearest neighbours method* computes a distance metric between a test data item and every other item in the data set (Witten and Frank, 2002). The distance metric is determined based on the difference in values for each of the item's features compared with each other data point. In most cases the distance metric is the Euclidean distance, however other measures can be used, such as non-linear measures that exaggerate the weight given when items have features with close values (Larose and Larose, 2014). At the start of the test, a value of k is selected by the analyst; once a distance metric has been calculated for each item, the k items with the lowest distance metric are considered. The test point is then classified by using a simple vote to determine the class that most commonly occurs in the nearest neighbours. If the vote produces a tie, then a further test is performed to find which of the tying neighbours are closer in order to provide a classification. The process is repeated for every unclassified item.

A problem when applying the k-nearest neighbours method to large data sets, especially for data where there are large numbers of features, is the computational effort to calculate the distance to every other point. However this problem occurs only where computational resources are scarce, or if results are required in a short period of time. Another problem is that in computing the distance metric there is a need to rescale all features onto the same scale (for example all value will be between zero and one) to ensure that all features are considered equally. Furthermore the method cannot be applied reliably in where there are *Boolean data* (*i.e.* features of the data that have the value either *true* or *false*; in some cases these are represented as 1 and 0, or *yes* and *no*) (Binmore, 2008). The data on level crossings contain many Boolean features, for example the presence of advanced warning signs is presented as a Boolean value (*yes* there are advanced warning signs, or *no* there are not); it is not meaningful to have intermediate values for this feature. For this reason, the k-nearest neighbours method is not suitable for data describing the many varied features of real-world level crossings.

2.15.7 Selection of machine learning methods for this study

Of the methods described above, five of the seven appear to be suitable for classification of level crossings. Table 2.7 summarises the findings from the prior sections.

Table 2.7 Summary of machine learning methods applicable for this study

Machine learning method	Suitability for this study
Decision trees	Suitable for use.
Random forest	Suitable for use.
Naive Bayes	Not suitable due to the *zero-frequency problem*.
Artificial neural networks	Suitable for use.
Support vector machines	Suitable for use.
k-nearest neighbours	Not suitable due to the occurrence of binary values in the input.

Within the literature there are many examples of the various machine learning methods being applied, however there is no clear information describing which methods perform better than others for a given task, or for a particular type of data. There are a number of papers that apply different methods to the same data in an attempt to identify which methods perform better than others, for example Domingues *et al.* (2018) and Nazari *et al.* (2018), however the results between different studies do not always agree. Machine learning is as an emerging technology and it appears that consensus is yet to be reached on which methods are preferred for a particular application.

2.16 Prior contributions

This study continues the author's prior work investigating level crossing safety, the causes of collisions, and methods to improve warning devices at level crossings. This study also extends prior work using modern data analysis techniques to understand safety risk on the railway.

The initial work was undertaken for Queensland Transport in 1999 and culminated the creation of an SRPT that was adopted for use by Queensland Rail to assess level crossing safety risk (Hughes, 2002). The tool included a method for assessing the safety risk of a level crossing both in its current configuration and after proposed risk controls had been applied. A rudimentary validation of the tool was undertaken, however a rigorous test of the tool's predictive accuracy was not possible due to limits of the data that were available at the time. The tool was subsequently adopted by all Australian road and railway authorities as the All Level Crossing Assessment Model (ALCAM, 2007).

A subsequent study was undertaken into the effect of different classes of warning devices on safety risk at level crossings in Australia resulting in a conference publication (Hughes, 2012a), which was reformatted for publication in a trade journal (Hughes, 2012b). At this time there had been an abiding belief in the concept of a hierarchy of controls (refer to Section 2.3.2) that assumes that active warning devices correlate with few collisions per road user traverse than passive devices do. Further it had been assumed that the provision of some form of road barrier at level crossings – usually a boom arm – correlates with a further reduction in collisions. At that time these assumptions had not been proven, again largely due to the absence of data on the number of road user traverses at level crossings. Nevertheless there was a general desire amongst road and railway authorities to provide active warning at all level crossings if possible.

A barrier to the widespread provision of active warning devices is the high cost of the equipment involved. Road users come to trust the information provided by level crossing warning devices (such as flashing lights) and believe that the absence of a warning necessarily indicates that no train is approaching. This approach is at odds with the general design philosophy of safety-critical information systems: usually systems are designed so that an active indication is required to show that a hazard is *not* present, which means that, if a warning device fails completely, users will assume that a hazardous state exists. However if a level crossing warning light assembly were to fail completely then it would look exactly as though it were safe for road users to proceed onto the level crossing. This inverted design of the warning means that high demands are placed on the reliability of level crossing warning devices: the devices must be tolerant to a large range of failures, such as power failures, and still be able to provide warnings. Furthermore the devices are situated next to railways and roads and have to be designed to be robust against a wide range of environmental conditions. These demands result in a high cost for any warning devices that acts as an impediment to the widespread installation of active warning devices.

Emerging technology has led to new designs of warning device that use features such as light-emitting diodes assemblies for providing warnings, solar power instead of wired electrical supplies, and wireless communications. These modern devices allow for substantial reductions in the cost of warning devices, however in some cases the low-cost devices cannot achieve the high reliability of prior technologies. In 2013, the author undertook a study in conjunction with researchers at the Queensland University of Technology to investigate whether there would be a net safety benefit by the widespread introduction of low-cost active warning devices, even if such devices could not achieve the same level of reliability as prior technologies (Wullems *et al.*, 2013). At around the same time, the author took over the directorship of ITS Innovations, a company that designed and manufactured low-cost level crossing warning devices. The company's product used a number of innovative design features that allowed them to operate for long periods without any maintenance or external power supply (Hughes, 2014).

A significant contribution of the author's prior work was a study undertaken in conjunction with a researcher at Imperial College London that used the newly published data from Network Rail (2017) to prove the long-standing assumption that active warning devices corelate with fewer collisions per road user traverse than passive devices do (Evans and Hughes, 2019). Another important contribution of this work was to consider the effect of level crossings on road traffic flow: the study derives a formula that shows that the total delay to road users as a result of a level crossing activation is proportional to the number of road users; the number of train traverses; and the square of the time the warning operates for each train approach.

Concurrent with this most recent work, the author undertook a number of studies to investigate how emerging computational methods can be applied to large and diverse data sets, and in particular whether application of these methods can be applied to improve the overall safety of the railway system (Van Gulijk *et al.*, 2018). These investigations have looked to identify efficient methods for collecting data from diverse sources to identify hazards arising from train drivers responding incorrectly to stop signals (Rashidy *et al.*, 2018), as well looking at more general methods for handling the very large volumes of data that would be encountered if a large number of diverse data sources were used for similar studies (Van Gulijk *et al.* 2016). The study of the use of large and diverse data sources has included a number of studies of how information can be extracted from text data to support railway safety. One study has focussed on identifying patterns in hazard reports written by railway infrastructure workers (Hughes *et al.*, 2016), and another identified trends in accident reports written in three different languages collected by the Swiss Federal Office of Transport (Hughes *et al.*, 2019). These studies of modern approaches to data analysis have informed the methods used in this study. It is noted that this study does not make use of any information derived from text data, however it is possible that future work could make use of text data – such as inspection reports or accident reports – to further improve safety at level crossings.

The work describes a study that is a corollary of these earlier studies and investigates in detail how volumes of vehicular road traffic affect rates of collisions at different classes of level crossings. In doing so, this work employs traditional statistical approaches as well as emerging machine learning methods for analysing the data. Some of the findings of this work form the basis for a joint submission with a professor of psychology to investigate how human behaviours affect safety at level crossings. The proposal for this subsequent study has already been submitted to Network Rail. These contributions are shown graphically in Figure 2.10.

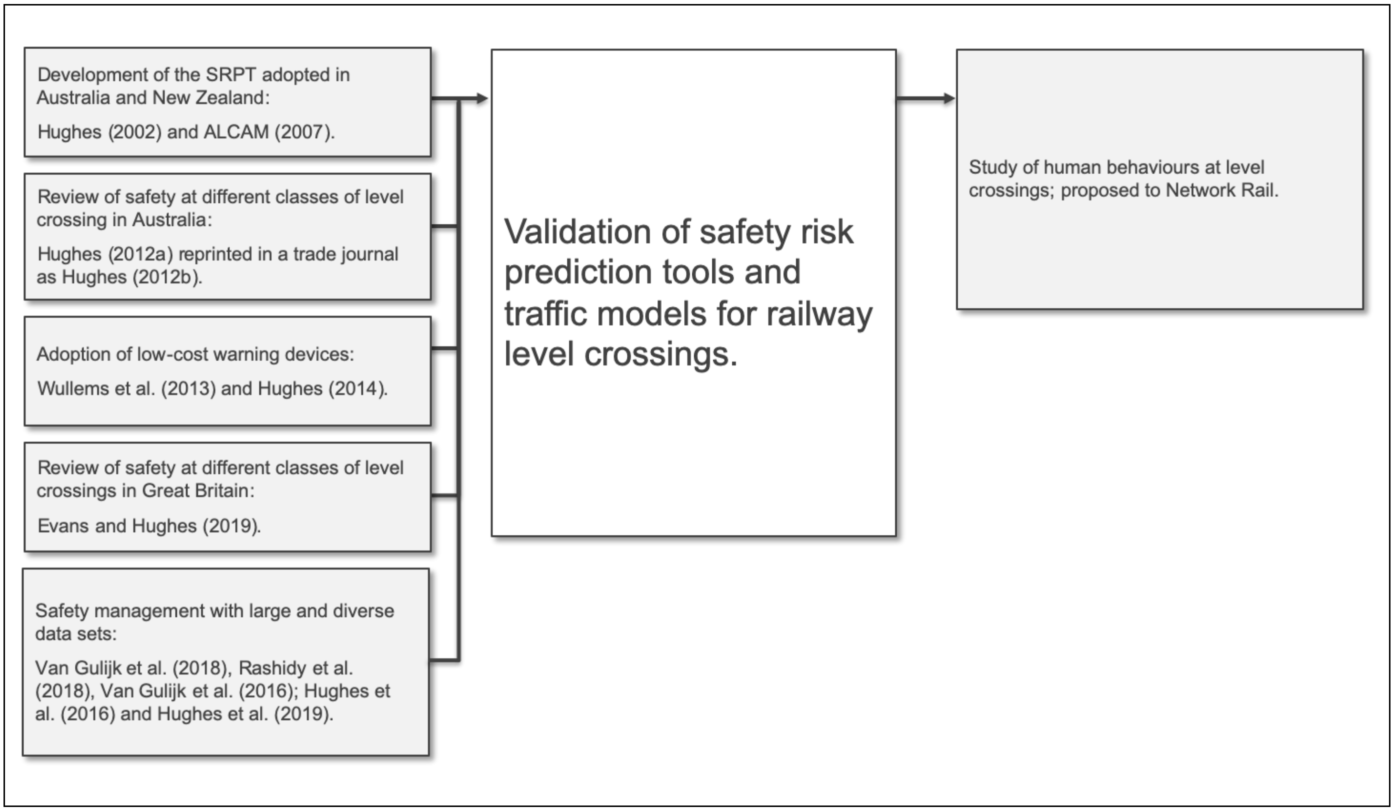


Figure 2.10: Graphical summary of prior contributions and proposed further study

2.17 Summary

There is a sizeable body of literature on level crossing safety. Overall, the literature show that there is no overarching theory of level crossing safety risk and in particular no consensus on the question how, or even whether, physical and operational characteristics of a specific level affect safety risk in general. Various researchers provide specific information regarding how individual characteristics have been observed to affect collision likelihood in specific cases. It should be noted that the various studies do not appear to disagree with each other, however no studies have been found that seek to confirm the results obtained by previous researchers. Whilst a number of the parts of a theory of level crossing safety seem to have been developed, no consistent overall theory has been developed.

Despite the lack of a general theory of level crossing safety risk, a number of road and railway authorities around the world have developed SRPTs. The SRPTs vary in the way that risk is calculated, and it can be expected that the various tools would give different risk predictions for the same level crossing. However such a test of correlation between the results of the various SRPTs is not possible with the information that is currently available: the details of the methods of calculation are not generally made available for a number of the tools and no calibration data can be found for any of the SRPTs.

A common feature amongst the various SRPTs is that they all use an underlying traffic model to determine how varying numbers of road vehicles and trains affect safety risk. Three different models were identified for which there is sufficient detail to perform verification of the model. These models are inconsistent with each other and would lead to different risk predictions for the same level crossing.

From the review of the literature it is clear that there is the potential for further study that would improve an understanding of the causes of safety risk at level crossings. Ideally there would be a complete, consistent and generally agreed-upon theory of the way in which physical and operational characteristics of a level crossing affect safety risk. Whilst various researchers have provided parts of the information that would be required for a full understanding, it is currently not clear whether such a goal is achievable. If it were, it is clear that there remains a significant amount of further work to realise that ambition.

The common feature of all the SRPTs is the underlying traffic models. Unlike the tools themselves, there is good information available on each of the models and the methods of calculation. In the pursuit of a general theory of level crossing safety risk, there is a clear opportunity for further research to provide a contribution by testing the degree to which the various traffic models correlate with observed collisions. The data from Network Rail that have recently been made available regarding road user volumes at level crossings are a valuable resource to testing the various traffic models.

Advancing information technology has allowed for data to be collected and analysed in ways that were not previously possible. A small number of researchers have identified ways in which these new analysis approaches could benefit management of safety risk on the railway. To date, however, these approaches have not been applied to managing level crossing safety. There is a clear opportunity for a study to review the traffic models that are used for level crossing safety risk prediction with a view to creating an overarching theory and, potentially, to use the emerging data analysis approaches that are currently being researched. There is an opportunity to use the new methods of machine learning, together with the newly available data, to establish whether it is possible to determine an SRPT that makes risk predictions that correlate with observed collisions. No studies can be found to date that have considered using machine learning methods in this way.

2.18 Contribution

The following contribution to current knowledge has been made in this chapter:

Contribution 2: This chapter has reviewed the literature that are available on level crossing safety and in particular SRPTs and the traffic models that underpin them. The review has identified the sources of data that are available for validating the SRPTs and methods of testing traffic models including emerging machine learning techniques.

Chapter 3: Investigation of traffic moment as a natural normaliser for level crossing collisions

3.1 Purpose and overview of the investigation

As discussed in Chapter 2, a number of SPRTs consider traffic moment to be a fundamental normaliser for collisions at level crossings. Despite this widespread use, no prior work can be found to test the validity of using traffic moment in this way. The apparent logic for such an assumption is straight-forward: a collision can occur only if both a road user and a train occupy are present on the level crossing at the same time. Therefore where there are either no road users or no trains then collisions cannot occur. As road use and train volumes increase, it can be expected that the number of opportunities for a collision to occur will increase. Indeed the product of the number of road user and the number of train traversals per day is sometimes called the *risk exposure* (for example see ALCAM, 2007). If it is assumed that collisions occur independently and randomly – for example as a result of inattention by road users in a way that is not affected by the actions of other road users – then the number of collisions at any level crossing will be in some proportion to the risk exposure. As such it can be expected that the number of collisions at a level crossing varies in accordance with the product of the number of road user and train traverses in a given period. Even if this core assumption is accepted – that collisions occur independently and randomly – it is not obvious that collisions vary with the product of road user and train volumes. For example if the distribution of road user arrivals and train arrivals at a level crossing is not constant during a day then it is not clear that an increase in traffic will lead to a linear increase is risk exposure. Perhaps road users are generally using a level crossing during the day with only a few arrivals during the night; whereas trains are generally using the level crossing at night with only a few arrivals during the day (such as may occur at a level crossing over a freight railway). In such a case it is not apparent that doubling the number of road users over a day will necessarily double the collision rate.

This chapter provides a theoretical consideration of the collision rates that could be expected for different road user and train volumes at a level crossing if arrivals of road users and trains occurred at random and road users took no action to avoid collision. Such a scenario is analogous to considering a road bridge over a railway line. With a bridge there is no opportunity for a collision, in this case road users will proceed over the bridge and trains will travel underneath without regard to the presence of the other. At some points in time a road user will be positioned on the bridge exactly over a train passing underneath. If the bridge had not existed, and trains and road users traversed the level crossing with no regard for each other, then such a situation would present a collision on the level crossing.

It is acknowledged that this model of a level crossing being equivalent to a bridge makes a number of simplifications: firstly at real-world level crossings, road users do respond approaching trains and usually avoid a collision. Another simplification is that road traffic flow cannot always be modelled by road users arriving at random, in some cases queuing can affect traffic flow. In spite of these simplifications, the number of SRPTs that use traffic moment as an underlying traffic model, it is worth testing whether this simplistic model has any validity.

Two studies were undertaken to determine whether traffic moment can be considered a fundamental measure of collision likelihood. Firstly a theoretical derivation was undertaken; secondly a Monte Carlo simulation was performed.

3.2 Theoretical derivation

The theoretical derivation considers trains and road vehicles passing over a level crossing at random. Figure 3.1a shows a train with length L_T and width W_T approaching a level crossing with speed S_T. Similarly the figure shows a road vehicle with length L_V, width W_V approaching at speed S_V. A collision will occur if both the train and road vehicle are occupying the conflict patch (P) at the same time; such a condition is shown in Figure 3.1b.

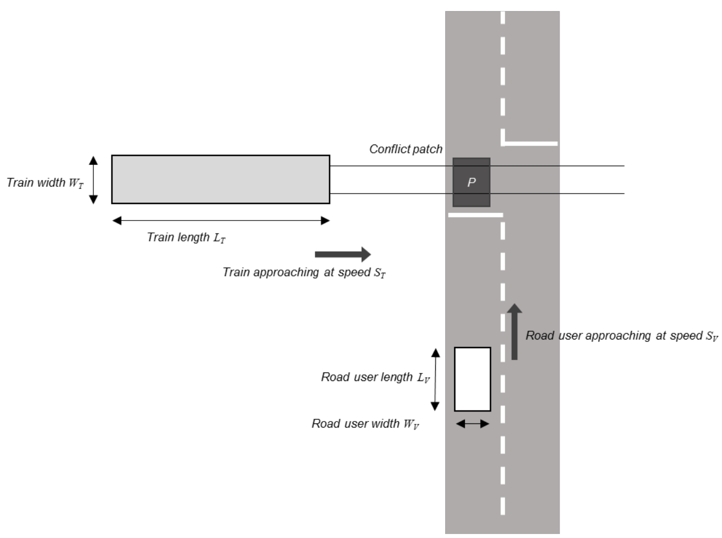


Figure 3.1a: A vehicular road user and a train approach a level crossing

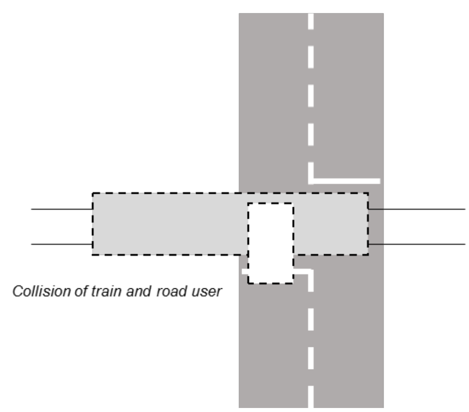


Figure 3.1b: Collision of train and road user

The conflict patch has a width equivalent to the width of a road vehicle. The entire length of the train must pass through the conflict patch. Therefore the amount of time the train spends in the conflict patch (t_T) is given by equation (1a):

$$t_T = \frac{W_V + L_T}{S_T}. \quad (1a)$$

By symmetry, Equation 1b shows the amount of time the road vehicle spends in the conflict patch (t_V).

$$t_V = \frac{W_T + L_V}{S_V}. \quad (1b)$$

For a day of duration D and T trains per day, the total proportion of a day that the conflict patch is occupied by trains, and therefore the probability of a train being on the level crossing at a randomly selected moment, is shown in Equation 2a.

$$P_{train} = \frac{Tt_T}{D} \quad (2a)$$

Similarly, for V road vehicles per day, the probability of a road vehicle being on the level crossing at a randomly selected moment is shown in Equation 2b.

$$P_{road\ vehicle} = \frac{Vt_V}{D} \quad (2b)$$

The probability of these random events coinciding is the product of the terms (2a) and (2b):

$$P_{collision} = \frac{Tt_T}{D} \times \frac{Vt_V}{D} \quad (3)$$

Rearranging gives:

$$P_{collision} = VT\frac{t_T \times t_V}{D^2} \quad (4)$$

Substituting from Equations 1a and 1b gives:

$$P_{collision} = VT\frac{\frac{W_T+L_V}{S_V}\times\frac{W_V+L_T}{S_T}}{D^2} \quad (5)$$

The result in Equation (5) shows that $P_{collision}$ is indeed proportional to VT (the traffic moment).

In addition to this primary finding, the result in Equation (5) also shows that $P_{collision}$ increases if:

- the length or width of either the train or the road vehicle increases, or
- *S* reduces, *i.e.* the vehicles travel more slowly over the conflict area.

The first of these points accords with intuition: that random collisions are more likely between objects of large size, simply because there is more area to collide with it is more likely that a collision will occur. The second point is less intuitive: collisions at level crossings could be expected to reduce if the speed of both road users and trains were increased. Taking this to the extreme would suggest that increasing the road speed of traffic approaching a level crossing to extreme levels could result in ever safer level crossings. Clearly such a finding is nonsensical: human drivers have a finite capacity for observing the route ahead and therefore limited ability to drive safely at increased speeds. Rather this second finding is a mathematical consequence of the idealised model that has been used and not a sound basis for road safety programmes.

It should also be noted that this simple model also disregards simultaneous traverses which can occur for one of two reasons, either by:

- two road vehicles arriving from different directions on a bi-directional road, or
- two road vehicles arriving from the same direction on a road with multiple carriageways.

Furthermore this derivation does not consider other effects such as queuing that could lead to road vehicles being stationary in the conflict patch.

3.3 Simulation method

Whilst the simplification in the previous section confirms the intuition that traffic moment is proportional to the number of collisions at a level crossing, the method of derivation may be overly simplistic in that it considers the rate of arrival of road vehicles and trains to be constant throughout the day. An alternative method to analyse the relationship between expected collision rate and traffic volumes is to use simulation. During a simulation the rate of traffic arrival can be varied during a day. In this way, the simulation can consider cases such as train arrivals that occur predominately during the night (perhaps as may occur on a freight line) and road vehicle arrivals that occur predominately during the day. Another advantage to performing a simulation is that it provides a second method of analysis that could act as a cross-check of the results of the mathematical derivation.

A Monte Carlo simulation was programmed for various road user and train volumes. The Monte Carlo method is a technique for risk estimation that uses a model of a system to calculate the outcome given a particular set of inputs. In this case the system is a level crossing; the inputs are the positions and speeds of approaching trains and road vehicles; the outcome is a count of how many collisions occur given these input values.

The model is recalculated a large number of times using different input values, which are randomly selected from a probability distribution. New input values are selected for each recalculation of the model. The results of many recalculations are aggregated to obtain an overall risk estimate.

Different simulations were run to allow for different total numbers of road users traversing the level crossing in a day. Twenty-one different values were used for the total number of road users in a day (V) in a range evenly distributed between zero road users per day to 5000 road users per day: therefore between each simulation an additional 250 road users per day were added. Twenty-five different values were used for the number of trains per day (T), distributed evenly between zero and 72 trains per day; a difference of three trains per day between each simulation. A simulation was run for each value of V against each value of T.

During a simulation, a day was divided in 24 one-hour periods. For each period, the expected number of road users and trains was calculated. Road users and trains were not considered to arrive at a constant rate during each one-hour period, rather arrivals occurred at random. The probability of an arrival was calculated to achieve the expected number of arrivals for each one-hour period. Arrivals were determined using a random number generator. Trains arrived singly at random, once a train was clear of the level crossing, the next train could not arrive within 20 seconds. Similarly, road users arrived singly; a new road user could not arrive within one second of the previous road user clearing the level crossing.

The length of the road vehicles used in the simulation were selected from a uniform distribution between 2 metres – which would be typical for a motorcycle – and

18 metres which is the longest vehicle permitted on public roads in the UK. The speeds of road vehicles were selected from a uniform distribution between 5 miles per hour and 60 miles per hour. Similarly the lengths of trains was selected randomly from a uniform distribution between 15 metres – representing a light engine – and 234 metres which would correspond to an inter-city passenger service. The train speeds were randomly selected between 5 miles per hour and 125 miles per hour.

A collision was recorded if a road user and a train were occupying the level crossing at the same time. It is possible for more than one road user to collide with the same train. The number of collisions that occurred in a day were summed.

To consider the effects of different distributions of road users and trains, five scenarios were considered for different proportions of road user arrivals and train arrivals for each hour of the day. The scenarios are discussed below.

Scenario 1: 'Double flat'

In this scenario, road user arrivals and train arrivals were constant throughout the day *i.e.* one twenty-fourth of the day's road users arrived during each hour of the day. Train arrivals were similarly distributed.

Scenario 2: 'Stott flat'

Arrivals of road users were varied over the day in accordance with the distribution proposed by Stott and shown in Table 2.4. Train arrivals were constant throughout the day, as for Scenario 1.

Scenario 3: 'Double Stott'

Both road user arrivals and train arrivals were varied in accordance with the distribution shown in Table 2.4.

Scenario 4: 'Double rising'

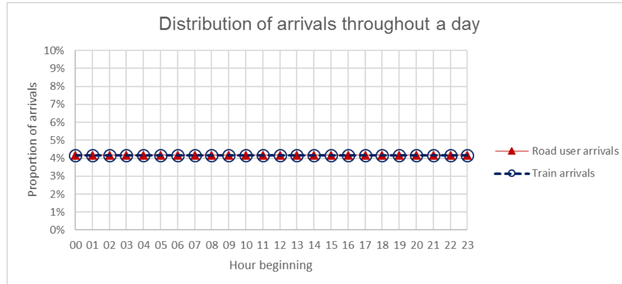
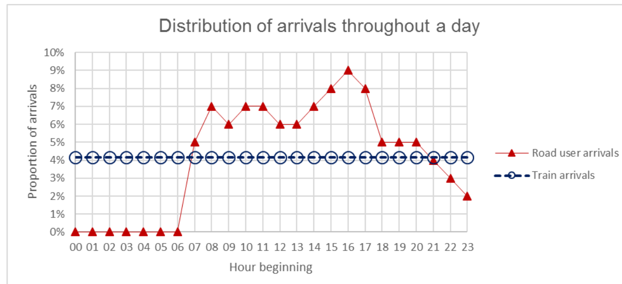
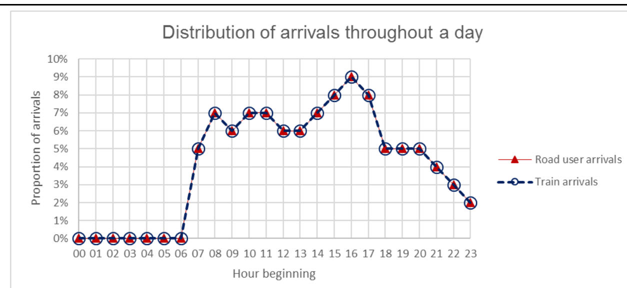
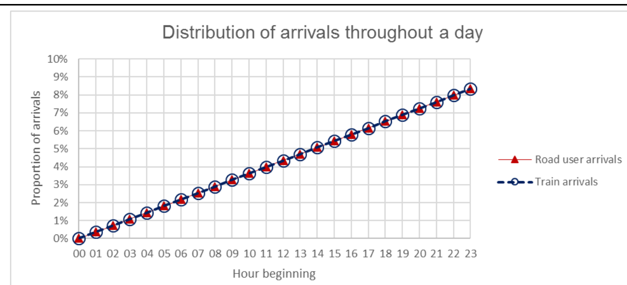
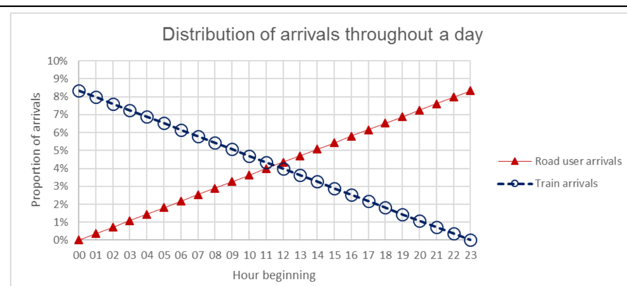
During the first hour of the day, road user arrivals were zero, the number of road users per hour was increased linearly throughout the day until the last hour of the day. Train arrival proportions were the same as road user arrivals.

Scenario 5: 'Rising falling'

Arrivals of road users were varied over the day as for Scenario 4, the number of train arrivals was the inverse of road user arrivals, *i.e.* train arrivals were at their maximum value during the first hour of the day and decreased linearly to zero in the last hour of the day.

Each of the scenarios is shown graphically in Table 3.1.

Table 3.1: Graphical representation of distributions over a day

Scenario 1: double flat	
Scenario 2: Stott flat	
Scenario 3: double Stott	
Scenario 4: double rising	
Scenario 5: rising falling	

For each scenario, simulations were run with different total numbers of road vehicles and trains arriving in the day. The total number of road users was varied between zero and 5000 road users per day in incremental steps of 250 road users per day (being 21 different values that were used in the simulations). Train volumes were varied between zero and 72 trains per day in incremental steps of three trains per day (25 different values). To obtain a large enough sample of statistical significance, each scenario was run for an equivalent of 10,000 days (approximately 27.4 years). The total number of trials in this study was therefore:

5 scenarios

× 21 values for road user volume

× 25 values for train volume

× 24 hourly periods

× 10,000 days

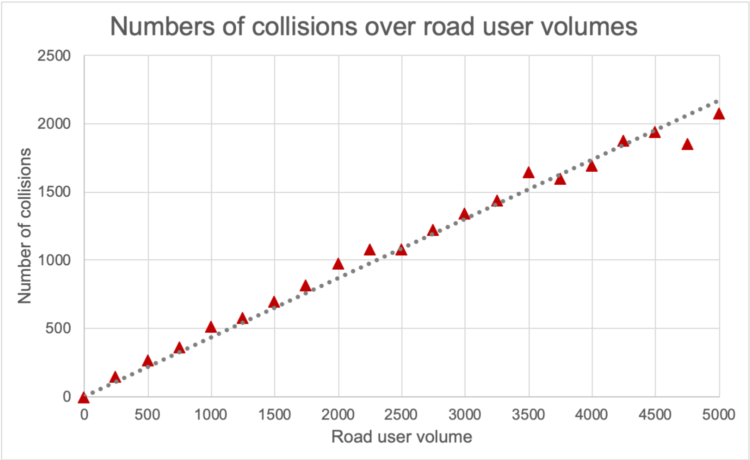
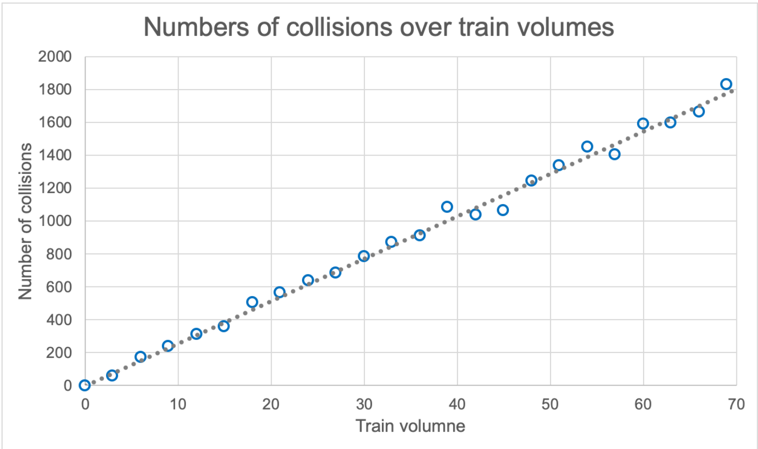
= 630 million random trials.

During the simulations, a collision was counted if a road vehicle and a train were occupying the level crossing at the same time.

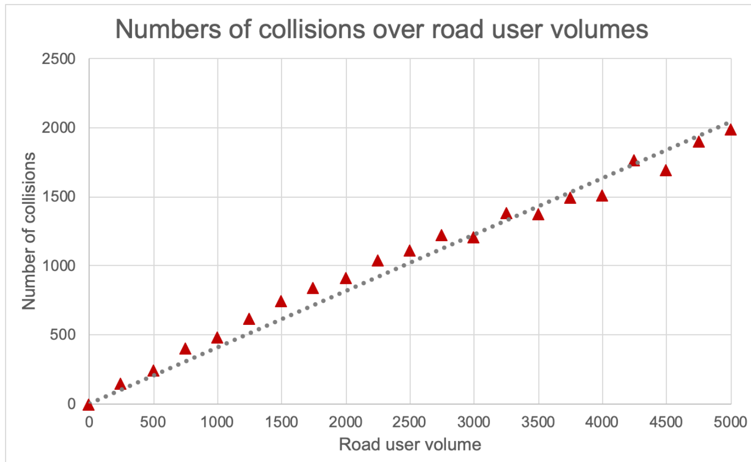
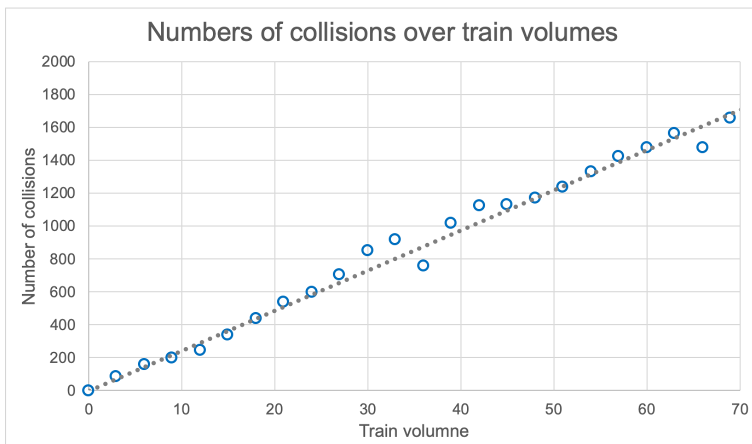
During the simulations it was assumed that road user will have an average length of six metres and travel at 40 km/h, therefore occupying the level crossing for an average of 0.54 seconds. Trains were considered to have an average length of 150 metres and travel at 100 km/h; giving a total occupation of 5.4 seconds.

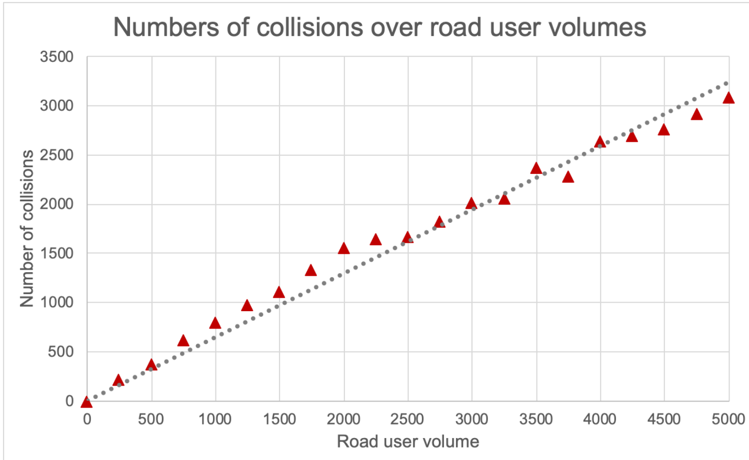
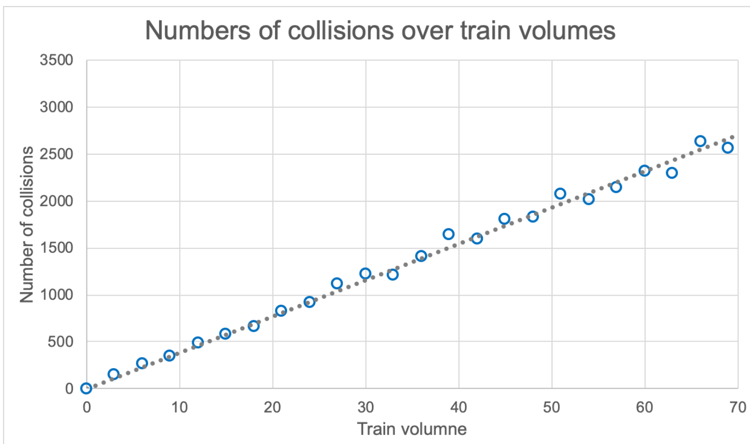
The number of collisions for each number of road user arrivals was summed for all trains arrivals for each scenario. Similarly the number of collisions for each number of train arrivals was summed for all road user arrivals. The null hypothesis assumed that collisions should vary linearly with increasing road user arrivals, and independently that they vary linearly with increasing train arrivals. To test this case, the numbers of collisions were plotted and a linear regression analysis was performed. The results are shown in Table 3.2.

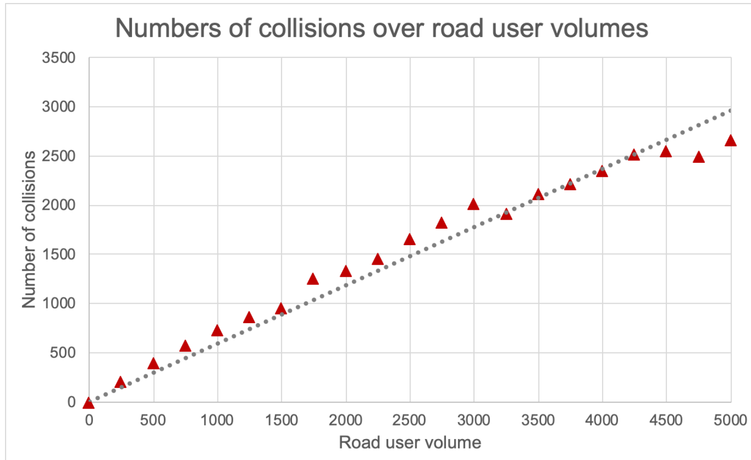
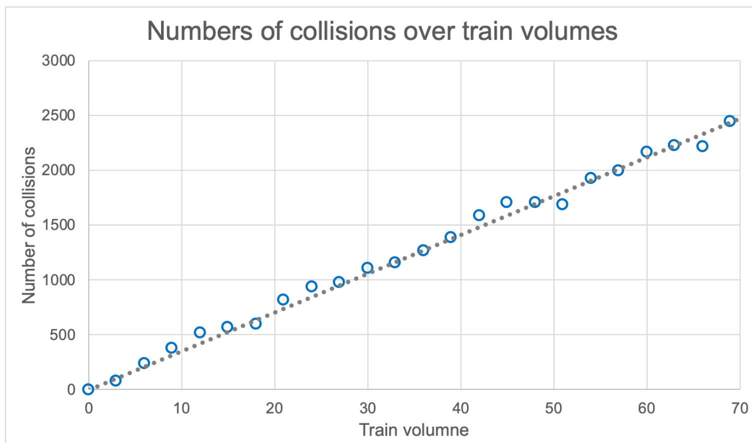
Table 3.2: Results of simulations

Scenario	Graphical results	R^2 value of linear regression comparison with traffic moment
Scenario 1 *Double flat*		0.9853
		0.9951

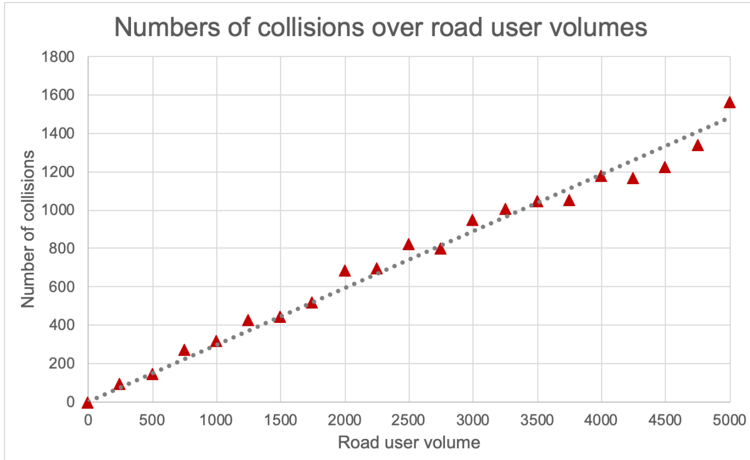
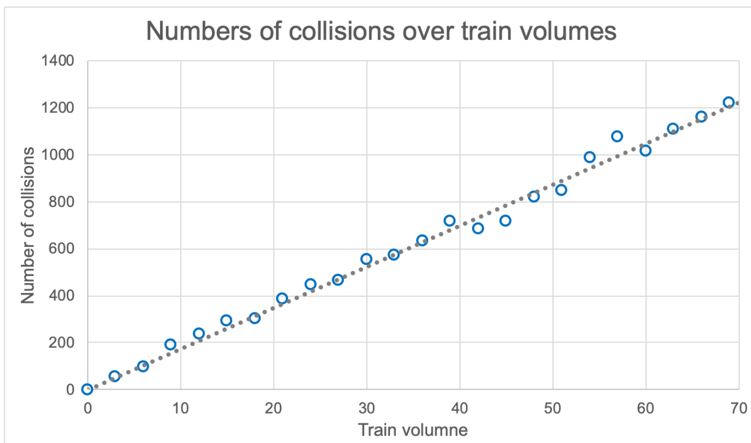
<table>
<tr><th>Scenario</th><th>Graphical results</th><th>R^2 value of linear regression comparison with traffic moment</th></tr>
<tr><td rowspan="2">Scenario 2:

Stott flat</td><td>
</td><td>0.9768</td></tr>
<tr><td>
</td><td>0.9867</td></tr>
</table>

Scenario	Graphical results	R^2 value of linear regression comparison with traffic moment
Scenario 3: Double Stott		0.9789
		0.9934

Scenario	Graphical results	R^2 value of linear regression comparison with traffic moment
Scenario 4: Double rising		0.9649
		0.9932

<table>
<tr><th>Scenario</th><th>Graphical results</th><th>R^2 value of linear regression comparison with traffic moment</th></tr>
<tr><td rowspan="2">Scenario 5:

Rising falling</td><td>
</td><td>0.9835</td></tr>
<tr><td>
</td><td>0.9918</td></tr>
</table>

3.4 Summary of results and discussion

The simulation results for each scenario show very strong linear correlation between collision rates and traffic rates for both road and rail traffic: in every case the R^2 values from the test vary between 0.9649 and 0.9951. Together the with results of the theoretical derivation, it can be concluded that for the simple case being tested the expected collision rates vary linearly with both road traffic and rail traffic. In combination the overall expected collision rate varies with the product of the road and rail traffic rates: *i.e.* the traffic moment.

This finding is perhaps intuitive: certainly the number of SRPTs that use traffic moment as the underlying traffic model suggests that when developing collision models for level crossings it appears natural to consider traffic moment as the measure for risk exposure. Despite this finding being apparently intuitive, no prior work had been found that provided any validation of this result.

Whilst this study has demonstrated the basic relationship between traffic moment and expected collision rate, the simplistic approach taken in the analysis and simulation is clearly incomplete. Firstly it is clear that road users do not proceed over level crossings without regard for approaching trains. Secondly, as noted by Stott, once a road user stops at a level crossing to yield to an approaching train, other road users who arrive behind will be unable to proceed onto the level crossing unless they bypass the road user who has stopped. Also, as noted above, the model used in this study considers only a single road carriageway with traffic approaching from a single direction. Where a road has more than one carriageway, or traffic approach from two directions, then it is possible that two road vehicles can be on the level crossing at the same time. In this case if one vehicle traverses the level crossing safely, then it is very likely that the other vehicle will also traverse safely. In these cases the traverses cannot be considered independent of each other and therefore one of the basic assumptions of the model is incomplete. Furthermore the simple model considers only *undersaturated* traffic flow, *i.e.* where vehicles are moving and are not being delayed by other traffic on the road. Saturation can occur, for example, when queuing occurs at a level crossing during the traverse of a train and the queue is unable to completely disperse before the arrival of the next train. Similarly, a common situation is that traffic may be queued over a level crossing as a consequence of a nearby road junction.

The hypothesis proposed by Stott addresses one of these simplifying assumptions – *viz.* that once one road vehicle has stopped to yield to a train then other road users behind are more likely to stop. However it may be ambitious to expect Stott's hypothesis to provide a fully accurate model since there are a number of cases that are not modelled in the hypothesis. For example Stott's hypothesis considers only collisions that arise as a result of road users breaching the level crossing holding point in the two seconds before a train arrives; whereas collisions can occur at level crossings when a road user has become stuck on a level crossing perhaps some minutes before the arrival of a train. Also, Stott's hypothesis considers only collisions where a train strikes a road user who is already occupying the level crossing, and not to cases where a road user strikes a train. Whilst there are few sources of information on the proportion of level crossing collisions that occur in this way; ATSB (2001) states that for 16% of all fatal level crossing collisions in Australia, the point of impact was the side of the train, and in a further 18% of fatal collisions the point of impact was not known. Stott's hypothesis also presumes that the first road user to stop behind the holding point at a level crossing will remain stopped and will prevent any further road users from breaching the holding point, however it is not clear that this will always be the case. For example if the first road user is a small vehicle (perhaps a moped) then subsequent vehicles can readily move around the first road user. Similarly if the second road user is a heavy vehicle then its failure to stop could easily cause a collision that shunts the first road user onto the tracks. Furthermore, Stott’s hypothesis is sensitive the distribution of road traffic throughout a day, however it is not clear that the daily distribution or road traffic proposed by Stott is correct for all, or even any, of the level crossings in Britain. As discussed in Chapter 2, Stott's hypothesis is a *predictive* model in that it was derived by logical reasoning, rather than empirical data collection. It is alluring to consider that it may be possible to construct a more detailed predictive model that contains elements from Stott's hypothesis and other factors to fully address the other factors that affect collision rate at a level crossing. However it is not clear that any such model could ever be developed: the Transportation Research Board (2000) note that in cases where saturation occurs it is not possible to create a general model of traffic flow since the actual behaviour of traffic depends on the characteristics of individual road layouts.

3.5 Conclusion

This chapter has provided theoretical rigour and validation to test the common notion that traffic moment can be used as an underlying traffic model for level crossing collisions. However it is noted that use of traffic moment requires some simplifying assumptions. Furthermore, the complexities of traffic movement, especially during saturated conditions, means that it is probably not realistic to expect that any single traffic model will be generally applicable for all level crossings.

The following chapters present an experiment to test the correlation between observed collision rates at level crossings in Britain and traffic models.

3.6 Contribution

The following contribution to current knowledge has been made in this chapter:

Contribution 3: The study described in this chapter has used two different methods to show that traffic moment is a valid normaliser of collisions at level crossings in simple cases where there is: unsaturated traffic; a single carriageway approach in each direction; no queuing from nearby roads; and road users take no care to avoid a collision.

Chapter 4: Method to test traffic models against observed collisions

The results of the theoretical derivation and simulation presented in Chapter 3 demonstrate that traffic moment is a valid traffic model in very simplistic conditions. The refinement proposed by Stott may provide a more accurate traffic model, however the model may be incomplete in that there may be a number of realistic scenarios that are not considered in Stott's hypothesis; such as the possibility of simultaneous train movements or the possibility that small vehicles stopped at a level crossing might not form an effective barrier to subsequent vehicles. This work therefore seeks to test the accuracy of the traffic models identified in Chapter 2 *viz*. traffic moment, Stott's hypothesis, and the Peabody Dimmick model – against observed collision rates on the GB railways.

This chapter describes the process of data collection and correlating observed collision data with level crossings; the process of calculating collision rates and introduces a meaningful unit for comparing the predicted collision rates with observed collisions. The chapter also investigates the statistical methods that can be applied for level crossing collision analysis.

4.1 Data collection

In order to test the traffic models, data are required on road and rail traffic volumes and observed collisions at individual level crossings. Network Rail publish data on level crossings in Britain. Section 4.1.1 describes the data that were collected for this study; and Section 4.1.2 describes the possible effects that errors in the source data may have on the results of the study.

4.1.1 Source data

The version of the data downloaded for this study (Network Rail, 2017) contains data on 6510 level crossings and describes, *inter alia*:

- a unique identification number for each level crossing (LX ID);
- the warning devices present at each level crossing;
- the number of trains traversing each level crossing per day; and
- the numbers of road users per day in two categories:
    - pedestrians and cyclists, and
    - vehicles.

Regarding the number of road users, Evans and Hughes (2019) note that the counts of traverses:

> *"are estimated from counts of users observed over short periods of time and grossed up to a full day. Some crossings are described as having "infrequent" use by pedestrians or motor vehicles or both. Such use is assumed for numerical purposes to be 0.5 traverses per day, but the precise assumption does not make much difference to the conclusions."*

Regarding the use of allocated 0.5 traverses per day and the potential for this to cause inaccuracy in the over result, they go on to note:

> *"Some of the data for individual crossings are likely to have large standard errors, but this paper assumes that the counts are reasonably accurate when taken over large groups of crossings."*

The data from Network Rail also contain some information on collisions in three categories: near misses; incidents; and accidents. In principle these data could be used to determine the number of collisions that have occurred at each level crossing, however no definition is provided to describe the meanings of these categories. It is not clear what differentiates a *near miss* from an *incident* from an *accident*. However data on all accidents on Britain's railway are collected in a central database, the *Safety Management Intelligence System* (SMIS) which is administered by the Rail Safety and Standards Board (RSSB). Data from SMIS are not generally available, but were supplied by RSSB on request. Data was obtained on all collisions between trains and road users at level crossings between 1996 and 2016 inclusive (being data for 21 full years). The data from SMIS detail:

- SMIS reference number;
- event date;
- level crossing category, although this is not coded in the same way as in the Network Rail level crossing data;
- the name of the level crossing;
- a textual description of the event;
- additional text data on the location of the level crossing including territory and location description; and
- in some cases, the level crossing identification number that corresponds with the LX ID in the Network Rail data.

Not all of these data were provided for every collision event: for a small number of events only the SMIS reference number and date were provided. Data on 491 events were provided, of these, 183 related to collisions with pedestrians which are out of scope of this study. Of the remaining events, eight were determined by RSSB to be self-harm which are also out of scope, leaving 300 collision events to be used in this study. Out of these events, an LX ID number was provided for 213 events. Where these data were provided they were assumed to be correct. For the remaining 87 events, an analysis was undertaken to determine if it was possible to correlate the event with a level crossing listed in the data from Network Rail. The correlation process is described in the next section. A summary of the collision data used in the analysis is provided in Appendix A.

This study has used data from different sources in a way that has not previously been done and the combination of data from different sources is a contribution of this study.

4.1.2 Effects of errors in the source data

The data for this study were provided external sources and it has therefore not been possible to confirm the accuracy of the data. The effects of errors in the road user count data and the collision data are considered below.

Firstly, considering road user count data, if there were significant errors in these data, there could be an effect on the correctness of the results from this study. Two cases need to be considered: *systematic errors* and *random errors* in the count data. If there were systematic errors in the data, the effect would be that the distribution of collisions over road user volumes would be incorrect. In such cases it is possible that even in cases where there is a correlation between the observed collision and the rates predicted by traffic models then the correlation may not be found as a result of the incorrect data on observed collision rates.

The data on the number of road user traverses were collected by Network Rail. During this process, there were some level crossings where no road users were observed during the counting period, in these cases the number of road users per day is recorded by Network Rail as *infrequent*. The presence of such level crossings in the source data demonstrate that there is no systematic effect that causes the counts to be too high at every level crossing in the data set. The remaining possibility for a systematic bias in the data would be an effect that causes the data to be tool low in each case. Again for level crossings where there are a small number of road user traverses per day, it is not clear how such an error could occur: it can reasonably be expected that staff performing road user counts are able to record small numbers of road users without error. If the data on road user counts is collected manually, it is possible that where there are large number of road users traversing a level crossing then there will be errors in the data. However it is not clear that the errors would systematically result in the counts being too low, rather it can more reasonably be expected that there would be random errors.

To some degree it is inevitable that there will be random errors in the data. The data have been collected from observations of the numbers of road users traversing each level crossing (Evans and Hughes, 2019): naturally there will be some fluctuation in the numbers of road users who traverse a level crossing each day. It is possible that the effects of some road user counts being too high could, to some degree, balance the effects on the overall analysis of some road user counts being too low. Clearly errors in the source data will inevitably lead to inaccuracies in the results of the analysis, however it is possible that small random errors may not skew the results the point that no correlation is found where a correlation does, in fact, exist.

Secondly, considering data on collisions at level crossings; if there were errors in these data, then there will be errors in the results of this study. Given that collisions are rare events, even small errors in the source data could cause a significant effect on the results of this study. The data provided by Network Rail have three categories of events at level crossing: *near misses*, *incidents*, and *accidents*. However there is no definition of any of these terms; for the purposes of this study it has been assumed that the data for accidents refers to collisions between road users and trains. Because of the uncertainty about the meaning of these data, another source of collision data was sought for this study. Data on collisions stored in the SMIS database were provided by RSSB.

Unlike the data from Network Rail, the data from RSSB are well defined. In the large majority of cases the collisions data include a text narrative of the event: it is inconceivable that the text narratives are fabricated and it is therefore not possible that there is over-reporting of collisions. Similarly it is unlikely that there is under-reporting of collisions at level crossings: any collision between a train and an object is a serious event and there are robust reporting processes in place for train drivers to report any type of collisions. It is extremely unlikely that a collision could occur without a train driver being aware, or without the driver reporting the event.

The data collected by RSSB are used for a number of purposes including directing investment for safety programmes and providing safety performance reports stipulated by legislation or regulatory requirements. If there were errors in the data collected and managed by RSSB then there would be a serious impact on the railway that would extend far beyond the scope of this study. The safe operation of the railway relies to a large degree on the data collected by RSSB to be accurate. As such it cannot reasonably be imagined that there are significant errors in the source data.

4.2 Correlation of collision event data with level crossings

For the 87 events where an LX ID number had not been provided, a systematic process was followed to attempt to identify the level crossing where the event had occurred and correlate this with data in Network Rail's spreadsheet of level crossings. Firstly, where a level crossing name had been provided in the event description, the Network Rail data were searched to see if a matching name existed. Level crossing names can be unique and where a name is provided that matches a single level crossing in the Network Rail data, then it is possible to obtain a good degree of confidence that the corresponding level crossing has been found. Examples of level crossing names include: *Sawbridgeworth Station*, *Bragg Marsh (Meldon Quarry)*, and *Munllyn*. In other cases the level crossing name was not given in the data field, however a description of the level crossing was often provided in the event narrative description. In some cases the narrative description provided a name that could be matched to the Network Rail data, however in some cases, even though a name was provided, a match could not be found. The SMIS narratives sometimes contained a description of the level crossing's location such as:

- *King's Lynn Service reported that he had struck a car at Coles Harbour LC approximately 3.5 miles north of Littelport Station*
- *Tan Lan crossing (UWG), approximately one mile on the approach to Llanrwst North*
- *at Johnstown, between Wrexham General and Ruabon*
- *Lowfield Farm UWC on the down Hull line*

Using these data, corresponding locations were sought using an online mapping service. Where a level crossing was found that appeared to match the description, the geographical location of the level crossing was extracted from the map as latitude and longitude. These location data were then compared with the location data provided by Network Rail to identify nearby level crossings. In some cases exact matches were found and the LX ID number for the corresponding level crossing was assigned to the event. In a small number of cases, additional information on level crossings was obtained from the ABC Railway Guide (2019) which is a publicly available data source that contains information on level crossing in Britain. Whilst the data in the ABC Railway Guide is, to a large degree, a reproduction of the information published by Network Rail, it does also contain photographs of many level crossings which can be matched against information from the online map. A further data source that was provided for this work was the ALCRM database from Network Rail. This data source is not publicly available and was made available only for the purpose of this study.

The data published by Network Rail contains information on current level crossings, whereas the ALCRM database contains data on *all* level crossings including those that have been closed. Consequently the ALCRM database contains data on many more level crossings that are in the publicly available data set. In a small number of cases it was possible to correlate the event described in the SMIS data with a level crossing in the ALCRM database that is not listed in the publicly available data; which indicates a level crossing that existed in the past (and at the time of the collision) but is no longer present. It is not uncommon for level crossings to be closed in response to collisions and therefore the reason the level crossing is no longer listed in the public data is because of the collision. Since collection of road user volumes at level crossings is a relatively recent activity, level crossings that are no longer in operation do not have road user volume data, which is necessary for this study. Consequently these events were not considered further in this study.

The process of allocating an LX ID to an event required care and cross-checking to avoid misallocation. For some events it was necessary to use all available sources of data to obtain sufficient confidence in the allocation of a level crossing.

During the work to correlate collisions with level crossings, four of the events in the SMIS data contained only an event number and a date: no further information was provided and it was therefore simply not possible to correlate these events with any level crossing. Similarly these events were excluded from further analysis. These events were among the earliest in the dataset with all occurring prior to November 1997; in is notable that, in general, the level of detail in the SMIS records appears to increase over time. One event had a description of the level crossing location which could be located on the online maps: the website's *satellite view* allows a level crossing to be seen at the location, however no nearby level crossings could be found in either the publicly available data nor the ALCRM database. It is possible that the omission of this level crossing from Network Rail's data is a consequence of a clerical error; alternatively it is also possible that the level crossing is not a formal, gazetted level crossing, but is rather an informal crossing point that has been established simply as a result of road users frequently traversing the rail at this point. In either case, road user volume data is not available and, again, the event could not be included in further analysis.

One of events that had been assigned to a level crossing by RSSB described a level crossing that could not be found in either the publicly available data from Network Rail nor in the ALCRM database (LX ID 32). However a corresponding level crossing could be found in the ABC Railway Guide which also listed road user volumes and therefore these data were used for the study. One collision occurred between a train and a motorcycle on a pedestrian footpath level crossing (LX ID 5988), since the vehicular road volume given in the data for this level crossing is zero, it was also excluded from the analysis.

Following the work to allocate collisions to level crossing, 284 events remained that could be used for further analysis.

4.3 Observation on collision data from SMIS and Network Rail

As well as containing details of level crossings, the spreadsheet provided by Network Rail also contains data on *accidents*, *incidents*, and *near misses*; however no definition is given for the meanings of these terms. It is assumed that the term *accident* refers to a collision between a train and a road user, however the data in the spreadsheet do not state whether the collision occurred with a *vehicular* road user or a *pedestrian*. For completeness, the analysis undertaken in this study used both sources of collision data: SMIS data and the accident data from Network Rail's spreadsheet. Ideally, it would be expected that the collision data from the different sources would be in exact agreement. Examination of the data identifies that this is not the case: there are significant differences between the collision data. For example, in the Network Rail data there is a level crossing that has a recorded accident history of five accidents, whereas the same level crossing in the SMIS data shows no collisions. Similarly there are three level crossings in the Network Rail data which are reported as having had four accidents, whereas no collisions are reported in the SMIS data for these same level crossings.

Table 4.1 provides a correlation matrix of the number of collisions recorded for each level crossing in each data source, the cells of the matrix show the number of level crossings with each number of recorded collisions. For example, it can be seen that there are 3358 level crossings that have no collisions recorded in either data sources; similarly there are 136 level crossings for which the SMIS data shows there have been no collisions whereas the Network Rail data show there has been one accident.

Table 4.1: Number of level crossings by collision counts from the Network Rail and SMIS data

		Number of accidents recorded by Network Rail					
		0	**1**	**2**	**3**	**4**	**5**
Number of collisions recorded in SMIS	**0**	**3358**	136	12	4	3	1
	1	181	**8**	1	0	0	0
	2	25	1	**0**	0	0	0
	3	6	2	0	**0**	0	0
	4	2	0	0	0	**0**	0
	5	1	1	0	0	0	**0**

The main diagonal in Table 4.1 is highlighted to show the only cells that would contain non-zero values if the data were perfectly correlated. The presence of non-zero values in unshaded cells shows the lack of correlation between the data sets. One reason for differences between the data sets may be that the data have been collected over a different time period. The data from SMIS contains a list of all collisions recorded in the database which, at the time of analysis was a period of 8107 days. Conversely the period of observation for accidents in the Network Rail data varied by level crossing from a minimum of 365 days to a maximum of 3,693 days. In all cases the observation period for accidents in the Network Rail data is less that the observation period for SMIS data. It could be expected that in some cases there would be more events recorded in the SMIS database than in the Network Rail spreadsheet. Such an effect would be seen in Table 4.1 as being non-zero values in the cells below the highlighted diagonal. There are a total of 219 level crossings where there are more events recorded in SMIS than in the Network Rail data. Conversely there are 157 level crossings where there are move events recorded in the Network Rail data than are recorded in SMIS. Consequently, the differences in the data cannot be readily attributed to differences in the observation periods.

Another reason for the differences in the data sets may arise from a difference in definitions, in particular events in the SMIS database are referred to as *collisions* whereas the Network Rail data refer to *accidents*. The SMIS data record any event where a train came into contact with a road user – either pedestrian or vehicular – at a level crossing, whereas there is no definition of an *accident* in the Network Rail. If there were a difference in definitions that somehow accounted for the differences in the data sets, the definition of an accident in the Network Rail data would somehow have to include events that are not defined as a collision in the SMIS data; *i.e.* there would have to be a type of accident that did not involve a train coming into contact with a road user, and any such accident would be outside the scope of this analysis.

When considering the source data and their use for risk analysis, it is notable that the SMIS data contain a larger amount of detail than the Network Rail data and in many cases include detailed descriptions of the events. Conversely the Network Rail source does not contain any additional data other than the count of events. The SMIS sources contains additional data that would allow the counts to be verified, whereas this is not the case with the Network Rail data.

Overall it can be seen that, within the GB railway industry, there is a lack of consistency in recording collisions at level crossings, it is possible that such differences could have a profound impact on national safety policy. Hypothetically it can be imagined that policymakers – for example RSSB – provide direction based on decisions made from analysing one set of data; and the operational railway – for example Network Rail – interpret policy and provide interventions on the railway based on another set of data. Such a miscommunication could lead to expensive safety interventions being applied where they are not required.

4.4 Unit of level crossing safety

Within the literature various measures are used for level crossing safety for example:

- RSSB record safety in units of injuries per year for four different types of injury to people: fatalities, major injuries, minor injuries, and shock and trauma (RSSB, 2016);
- Evans and Hughes (2019) measure safety in units of fatalities as a result of collisions per million road user traverses;
- Barić *et al.* (2018) report the number of accidents per level crossing with the area being studied; whereas
- the US Department of Transportation (2007) reports injuries per level crossing.

It is not clear that all of these measures are meaningful for comparing collision rates between level crossings. It may feel intuitive to measure safety in terms of the number and nature of injuries arising from a hazards such as fatalities and major injuries, however such a unit may not be meaningful for proactive management of safety. Whilst it is possible for road or railway authorities to provide controls to reduce the likelihood of collisions occurring, it is not clear that they can necessarily influence the outcomes of collisions in terms of the numbers or nature of injuries. Injuries resulting from any collision may depend on a number of factors such as:

- the number of people within a road vehicle;
- pre-existing health conditions of individuals;
- chance matters of timing of maybe a fraction of a second that may affect the exact parts of the train and the road vehicle that come into contact;
- the crashworthiness of the vehicles involved; and
- presence of hazards within a vehicle (such as unbound objects that are thrown through the passenger area) or absence of safety controls (such as air-bags, or containment systems to prevent fuel leaking after a collision).

These factors cannot be predicted nor controlled in advance by either the road nor railway authorities and must therefore be considered to random effects arising as a result of collisions. It is clearly not meaningful to measure the efficacy of safety management interventions in terms of random effects, rather safety should be measured in terms of factors that can be influenced by management interventions which.

Another consideration is that it is necessary to normalise the collision rate by some measure of exposure to collisions. As stated earlier, a collision at a level crossing requires the presence of at least one train for a road user to collide with. Again, it may appear intuitive to normalise the rate of collisions by the number of trains traversing a level crossing in a given time. Whilst the exact measure of exposure may be related to the number of trains, the exact measure more complicated. For instance where a collision occurs as a result of a road user striking the side of a train that is already occupying the level crossing (as described in ATSB, 2001), the measure of exposure is related not only the number of trains but also the length of the trains. Furthermore, where there are duplicated tracks it is possible for two trains to traverse a level crossing (often in opposite directions) at the same time. In cases where two trains are traversing a level crossing simultaneously then, compared to a single train, the amount of time in a day when collisions can occur is clearly not doubled. Conversely collisions have occurred at level crossings as a result of road users waiting for the first train to pass then moving into the level crossing unaware that a second train is approaching. The occurrence of simultaneous train traverses at a level crossing therefore has a complicating effect on the risk exposure which cannot necessarily be modelled by simply considering the overall exposure to be doubled. The data available on train movements in Britain do not describe the lengths of trains, nor do they describe situations where trains simultaneously traverse a level crossing. It is therefore not possible to provide a complete description of the risk exposure from the available data. The majority of train traverses at level crossings do not occur simultaneously with other trains, therefore a simplifying assumption for this study is that there are no simultaneous train traverses. Furthermore it is assumed that the risk of a collision occurs as the result of the presence of a train on a level crossing regardless of the length of a train. In effect the assumption is that all trains are of approximately similar length. The data provided by Network Rail describe the number of train traverses per day. Therefore the unit of risk exposure used in this study will be per train traverse of each level crossing for a given time (in this case the units will be *per day* as used by Network Rail).

The unit of level crossing safety that will be used in this study will therefore be: *collisions per train traverse per day*.

For this study, the purpose is to measure the effect of road traffic volume on level crossing safety risk; as such the independent variable is the measure of road traffic volume. The unit used in this study will be the data provided by Network Rail describing the estimated number of traverses of a level crossing each day by road vehicles.

4.5 Methods for statistical analysis

The purpose of the analysis is to determine the degree of correlation between observed collisions and traffic models used in the SRPTs. There are a number of standard tests that can be applied to test correlation between two sets of data. However not all tests can be meaningfully applied to all data sets and it is necessary to give consideration to the nature of the data in order to select an appropriate test or tests. One correlation test that is commonly used is the *chi-squared* test. However, an important caveat is that this test can be applied only when the data being tested are dimensionless: this point is clarified by a number of sources, for example Oliveira and Oliveira (2013) state: "*we can use the* [chi-squared test] *to investigate associations between two binary variables, between a binary variable and a categorical variable, and between two categorical variables*". In this case it is important to note that a *binary variable*, is a type of categorical variable that has only two values. Other sources that emphasise this point about the chi-squared test include:

- Crawford and Csomay (2015): "*with the chi-square tests, both the dependent and independent variables can be nominal data*".
- Hinton *et al.* (2014): "*a chi-square statistical test allows you to analyse frequency data".*
- Mackridge and Rowe (2018) take a blunt and didactic approach regarding the chi-squared test: "*you will use two categorical variables*".

Gunawardena1 (2011) provides advice on selecting statistical tests and provides a number of diagrams illustrating the choices to be made in selecting an appropriate test. The first choice depends on the nature of the data. Figure 4.1 shows an adaptation of one of the diagrams showing that the chi-squared test can be applied when the data being compared are *nominal*. Other graphics in the source show tests that can be applied for other types of data: the chi-squared test is not shown as a test that can be used for any data type other than nominal data.

The variable being analysed in this study is a continuous variable that has units of *collisions per train traverse per day* and therefore has dimensions of TIME^{-1}, as such it is clear that the chi-squared test would not be suitable for this study.

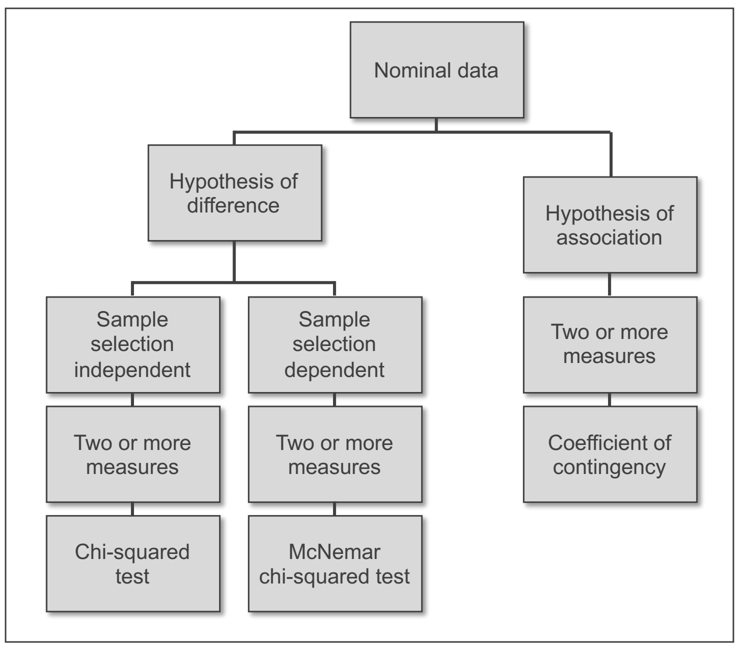


Figure 4.1: Selection of a statistical test for nominal data, adapted from Gunawardena1 (2011)

In some studies, it may appear *prima facie* that the chi-squared test has been applied to dimensioned data. However in these cases the test has not been applied to the source data, rather the source data have been binned into categories and the count of the numbers of entries in each category are used as the test variable in the chi-squared test. As such, this application of the chi-squared test uses the categorical data that are derived as a result of binning. In order to perform a test in this manner it is necessary to select appropriate bins for separating the data, and it is important to note that the selection of bins may affect the outcome of the study. In some cases there are natural categories to separate the data into, however for continuous variables the selection of bins is, to some degree, arbitrary which would result in an unpredictable effect on the outcome of the test. As such, use of the chi-squared test in this study could become cumbersome and would be unnecessary since there are other tests of correlation that work with dimensioned data and do not require arbitrary binning.

A further consideration is the sparse nature of the data that are used in the study. There are many level crossings where no collisions have been observed, as such a large proportion of the observation data have a zero value. Even if the data were to be binned, there would still be many bins that have zero observations, so the data would remain sparse; Renter *et al.* (2000) note that the chi-squared test is "*generally unreliable*" when applied to sparse data. Again, there are other statistical tests which do not have the same problem with sparse data.

When considering statistical tests it is useful to identify the data type of the variables under test using the scales of measurement developed by Stevens (1946), which classifies data into one of four types: *nominal*, *ordinal*, *interval*, and *ratio*. The variable under test is dimensioned as discussed. The test variable is also properly sequenced in that (for example) two collisions per train traverse per day can meaningfully be considered to be twice one collision per train traverse per day; furthermore there is a meaningful zero value. Using Stevens' classification, the test variable is therefore a *ratio* variable.

There are two further considerations when identifying a statistical test. The first is that the correlation test has to be able to test the degree of correlation between two non-parametric data sets. A number of the statistical tests test for *normality* of a data set, however the traffic models used in the SRPTs do not assume that collisions are normally distributed. Furthermore the Peabody Dimmick traffic model, being a descriptive model, is based on observed data and cannot be defined as a parametric model. Therefore any statistical test must allow for a test of correspondence between non-parametric data.

The second consideration is that − fortunately − level crossing collisions are rare: in fact most level crossings operate for many years without any collisions occurring. Consequently when analysing level crossing by the number of collisions (per train traverse per day) that have occurred, the data set contains a large number of level crossing where the result is zero. In fact the mean number of collisions per level crossing is close to zero. Conversely, since there are some level crossings that have a history of multiple collisions, there is a non-zero standard deviation in the data. The standard deviation is greater than the mean value leading to a situation where within one standard deviation of the mean are negative values for the rate of collisions which is clearly meaningless. The term *overdispersed* is applied to such data sets. A number of the statistical tests described in the literature assume that, even if the data are not normally distributed, they are nonetheless not overdispersed. The test to be applied in this study therefore needs to be robust against the effects of overdispersion in the data.

There is a substantial body of literature describing statistical tests and the applicability of each test. Literature on statistical tests were reviewed to identify a suitable candidate (Charan, 2010; McCrum-Gardner, 2008; Kanji, 2006). Within the literature there is a commonality in terms of the tests described: amongst the literature describing statistical tests in general it is possible to find a number of descriptions of the same test (for example the chi-squared test), however much of the literature fail to describe any tests that can be applied for overdispersed, non-parametric data. It is necessary to research specialist literature in order to identify tests that are suitable for over-dispersed and non-parametric data (for example Skovlund and Fenstad, 2001; Dean and Lundy, 2014). During the review a number of candidate tests were identified, however for various reasons these tests could not be applied. For example the *Anderson Darling* test, the *t-test*, and the *Shapiro Wilk* test can be applied only where the data are normally distributed.

One candidate, the *Mann-Whitney U* test, is a ranking test of *ordinal* data and therefore does not immediately appear to be relevant to the *ratio* data in the test variable. Nevertheless it is possible to create a set of ordinal (ranked) data from a set of ratio data. It would then be possible in principle to compare the rankings of the data in order to perform some test on the degree of correlation between predicted and observed collision rates. However, in practice this is not the case for overdispersed data. Of the 3742 level crossings considered in this study, most have no history of collisions. Therefore a ranked set of the data would have all of these level crossing ranked equally at the last positiin in the list. However, each of the traffic models predict some non-zero collision probability for any level crossing that has non-zero road or rail traffic volumes. Therefore any test that compares the rankings of the data sets will find that 95% of the values will have an incorrect ranking. It is therefore not meaningful to apply a ranking test on overdispersed data sets.

As a result of the review, one test was identified that is applicable for this analysis: the *Kolmogorov Smirnov* test (Aeron *et al*., 2011; Antoneli *et al*., 2018; Feigelson and Babu, 2013). The test considers the correlation of two non-parametric data sets by comparing the cumulative proportion of the test variable against the cumulative proportion of the comparison data set over the full range of values. If the data sets are completely correlated then there will be no difference in the values at any point in the range. The Kolmogorov Smirnov test considers the maximum difference between the cumulative values and, for a given significance level (also known as the *alpha value*, α) whether the two data sets can be assumed to be drawn from the same population. The Kolmogorov Smirnov test is insensitive to overdispersed data: where there are zeroes in the observed number of collisions, the cumulative value does not increase, however it is still possible to compare the cumulated value at that point on the curve with the cumulative number of predicted collisions. A further, serendipitous, property of the Kolmogorov Smirnov test is that it does not consider the absolute values of the variables being tested, merely whether the variables follow a similar shape over the range of values. This property is valuable for the study since the absolute collision rates vary significantly between classes of level crossing, and the traffic models do not necessarily correctly predict absolute numbers of collisions. In presenting his hypothesis, Stott proposed a general distribution of how collision rates would vary with numbers of road vehicles. The absolute numbers of collisions predicted by Stott's hypothesis depend on the value of the variable P_c. However Stott notes that the value of P_c used in his paper is not based on observation, instead it appears to be arbitrary as he notes: "*P_c must indeed be very small – of the order of 1 in 10^4 or less*". Conversely, the Peabody Dimmick model predicts collision rates that are based on observed collision rates in the United States in the 1930s (Faghri and Demetsky, 1986). Changes in road transport technology in the time since have significantly altered the overall safety of road transport. Consequently, whilst the overall effect of increasing road traffic numbers at a level crossing may vary in accordance with the predictions made by the Peabody Dimmick model, it cannot be expected that the absolute number of collisions observed would be the same as predicted by the model. Furthermore it cannot be expected that the number of collisions predicted by Peabody Dimmick would be consistent with Stott's (arbitrary) estimate of collision rates. The Kolmogorov Smirnov test is therefore useful as it provides a means to test the principles proposed by the traffic models without being sensitive to the absolute numbers of collisions.

The lack of a test of absolute values does not present a problem to this study. If it is found that a particular model correlates with observed collision rates, it will be possible to analyse the results to determine what the actual probability of a collision per traverse should be; in effect it would be possible to use the analysis to determine the correct value of Pc to use in Stott's hypothesis or any equivalent factor for the Peabody Dimmick model.

4.6 Differences in the curves of absolute and cumulative values

The Kolmogorov Smirnov test compares the *cumulative proportion* of the samples rather than the absolute values. As a result, all values are scaled to be in the range from zero to one (more usually written as 0% to 100%), therefore all curves on the graph always touch the origin at the bottom-left of the graph (0%, 0%) and take a path to the point (100%, 100%) at the top-right of the graph. Since *cumulative* values are used, the shapes of the curves when shown on a graph are not the same as the curves for the absolute predicted values. For example, Figure 2.6 shows the distribution of collisions hypothesised by Stott, showing that the hypothesised number of collisions initially rises as the volume of road vehicles increases, reaches a peak value and then falls. The same rising-then-falling shape cannot be seen in the values in the graphs in Chapter 6 since the cumulative value will never decrease. Instead the cumulative value of the curve shown in Figure 2.6 will initially rise quickly, then the rate of increase will reduce as low hypothesised numbers of collisions lead to only a small increase in the cumulative value.

4.7 Kolmogorov Smirnov test null hypothesis and interpretation of alpha values

The Kolmogorov Smirnov test is a test for correlation between two samples of data. The test assumes a null hypothesis that the samples are drawn from the same distribution, *i.e.* that there is correlation between the two samples. Obviously it would be very unlikely for any two samples to correlate exactly; natural variation in the data will almost certainly create differences between any two samples. Where the differences between the two samples are small, it is likely that the null hypothesis is valid.

Conversely, where the differences are large, it more likely that the data are not in fact correlated, and that some effect other than natural variation explains the differences between the samples; in such cases it is assumed that the data are taken from different distributions and therefore the null hypothesis should be rejected.

The *significance level* for the test, also known as the *alpha value* (α), is the probability that natural variation in the data leads to differences so large that the null hypothesis is rejected when it is in fact true. For example, an α value of 1% describes a 1% chance of rejecting the null hypothesis, when in fact the samples were drawn from the same distribution. Therefore an α value of 1% means that rejection of the null hypothesis is less likely than rejection with a higher α value, say, 20%. Consequently the α value of 1% requires more variation in the data than an α value of 20% before the null hypothesis is rejected. In general, the lower the α value, the more difference is required before rejection occurs. The Kolmogorov Smirnov test produces a value that is a measure of the cumulated differences between the samples. The test value is compared with a critical value to determine whether the null hypothesis can be rejected for a given α value. If the cumulated difference is less than the critical value, then the variation in the data is considered to be attributable to natural variation and the null hypothesis is not rejected. In summary, because the null hypothesis for the Kolmogorov Smirnov test is the assumption that the data are correlated, the α value is a measure of the likelihood that the null hypothesis is true, but has been falsely rejected. Consequently, the meaning of the α value is often counter-intuitive, and it can seem that more variance is tolerated when the α value is higher, which is not the case.

4.8 Summary of study method

The level crossings listed in the Network Rail data were categorised into six categories as used by Evans and Hughes, *viz*:

- passive vehicular public,
- passive vehicular private or staff,
- automatic vehicular public,
- automatic vehicular private or staff,
- railway-controlled vehicular public, and
- railway-controlled vehicular private or staff.

Collision data from the SMIS database were corresponded with each of the level crossings in the Network Rail database. Using the census data of road user and train traverses for each level crossing, a test variable of collisions per train traverse per day was calculated for each level crossing. As noted above, for the majority of level crossings, this value is zero. A Kolmogorov Smirnov test was performed to test the degree of correlation between the test variable and each of the three traffic models, being:

- traffic moment,
- Stott's hypothesis, and
- Peabody Dimmick's model.

As discussed, the data from Network Rail also include for each level crossing, counts of:

- near misses,
- incidents, and
- accidents.

Whilst these terms are not defined, the data themselves may nevertheless be meaningful in some way. Consequently the test described above was repeated using the data for *accidents* from the Network Rail data in place of the collision data from the SMIS database.

Overall there are six categories of level crossing, and two sets of data describing either accidents or collisions, making a total of twelve sets of observation data. Each of these observation data were compared against the three traffic models, making a total of 36 tests that were performed. The purpose of the tests is to determine the degree to which the observed accidents or collisions corresponds with the traffic models.

4.9 Contribution

The following contributions to current knowledge have been made in this chapter:

Contribution 4: This study has established *collisions per train traverse per day* as a unit that provides a meaningful way to compare collision rates between level crossings.

Contribution 5: A rigorous review was undertaken to identify a suitable test to allow a meaningful comparison of observed collision rates against traffic models. In particular the method of testing needs to be robust in cases where data are overdispersed. The *Kolmogorov Smirnov* test was identified as being appropriate for this analysis.

Chapter 5: Application of method, analysis results and interpretation

This chapter presents the results of the tests of correspondence between the predictions of the traffic models and the observed collision rates. A Kolmogorov Smirnov test was performed to determine the degree of correspondence between the predictions of various traffic models and observed collision rates. Six categories of level crossing were considered:

- passive vehicular public,
- passive vehicular private or staff,
- automatic vehicular public,
- automatic vehicular private or staff,
- railway-controlled vehicular public, and
- railway-controlled vehicular private or staff.

Two different sources of data were used for the observed collision rate data:

- collision data from the SMIS database, and
- accident data supplied by Network Rail.

The independent variable in the tests was the number of road users traversing each level crossing in a day, the dependent variable being tested was the predicted collisions per train traverse per day. Tests were carried out for the predictions made by the traffic models:

- traffic moment,
- Stott's hypothesis, and
- Peabody Dimmick's model.

Since there are six categories of level crossing, two sources of collision data, and three different traffic models, a total of $6 \times 2 \times 3 = 36$ tests were performed. The subsections below present the results for one of the level crossing categories and one source of observation data (giving 12 subsections), and show the results of tests with each of the three traffic models. Each subsection contains a graphical representation of the values calculated in the Kolmogorov Smirnov test. In the graphs:

- the observed collision or accident data are shown as orange squares (◆);
- traffic moment predictions are shown as unfilled blue circles (○);
- Stott's collision hypothesis predictions are shown as filled blue circles (●); and
- Peabody Dimmick collision predictions are shown as blue triangles (▲).

The graphs allow for an intuition of the correspondence between the observed collisions and accidents and the traffic model predictions: where the orange squares are close to any of the blue markers then it can be seen that the observations closely correspond to the traffic model. The Kolmogorov Smirnov test produces a single value which is a measure of how closely the two distributions correlate with each other over the entire range of the independent variable. The value derived from the Kolmogorov Smirnov test is compared against a *critical value* to determine whether the correlation is statistically significant for a given alpha (α) value. Kanji (2006) provides critical values for five α values, *viz.*: 1%, 5%, 10%, 15% and 20%. In each subsection a table is provided showing the value produced from the test for each of the traffic models when compared with the observed data; the critical values for each of the five α values; and a comparison of whether the test result suggests that the result is significant for each α value.

The results in the following sub-sections contain a table that compares the test result from the Kolmogorov Smirnov calculation with the critical value for different values of alpha to determine whether the null hypothesis (that the data are drawn from the same sample) should be rejected as discussed in Section 4.7. The critical values for the Kolmogorov Smirnov test are supplied in a table for cases where there are 35 or fewer test cases (for example, Kanji, 2006); for cases where there are more than 35 test cases (n) the critical values are calculated as shown in Table 5.1 (Massey, 1951):

Table 5.1: Calculation of critical values in the Kolmogorov Smirnov test

Alpha (α) value	**Calculation for critical value for different numbers of samples (n)**
1%	$\frac{1.63}{\sqrt{n}}$
5%	$\frac{1.36}{\sqrt{n}}$
10%	$\frac{1.22}{\sqrt{n}}$
15%	$\frac{1.14}{\sqrt{n}}$
20%	$\frac{1.07}{\sqrt{n}}$

The values in Table 5.1 show that rejection is more likely to occur for higher values of α, as discussed in Sections 4.5 and 4.7.

The figures in Sections 5.1 to 5.12 provide a graphical representation of the results of the Kolmogorov Smirnov tests. In each graph the cumulative proportion of collisions per train traverse per day is plotted over the proportion of road users at each level crossing. Since both the abscissa and ordinate are showing *proportional* values, the scales on each axis are a percentage value between 0% and 100%. The value on the ordinate axis is the *cumulative* proportion of collisions per train traverse per day and will therefore never decrease over the range of road user volumes: in cases where there are no observed collision, the cumulative value will remain unchanged; where collisions are observed the cumulative value will increase. It is for this reason that all of the values in the graphs can be seen to rise from the origin (0%, 0% in the bottom-left corner), to the value 100%, 100% (in the top-right corner).

Discussion of the results is provided in Sections 5.13 to 5.20.

5.1 Passive Vehicular Public - SMIS collision data

Results from analysis of the 112 level crossings categorised as *passive vehicular public*.

Graphical results

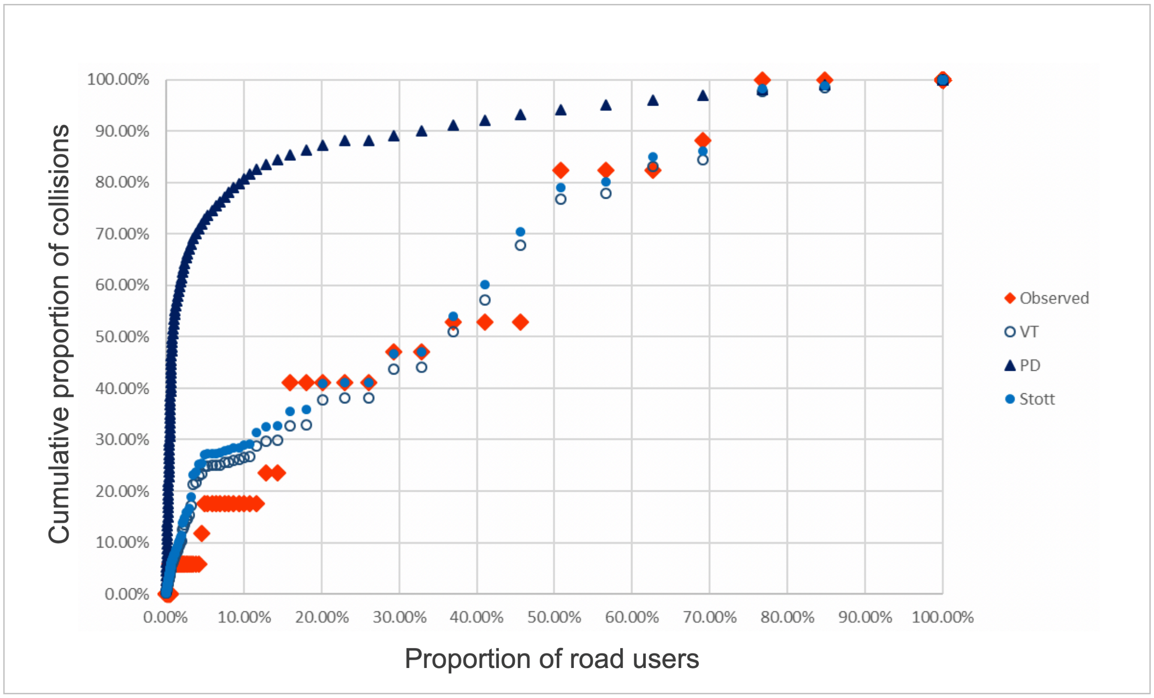


Figure 5.1: Observed collisions for passive vehicular level crossings using SMIS data

Passive vehicular level crossings on public roads are uncommon in Britain, and occur only on roads with few road users per day. The maximum value of V for level crossings in this class is 570 road users per day; compared to nearly 30,000 for railway-controlled vehicular public level crossings. For low values of V, the distribution of collisions predicted by Stott's hypothesis rises approximately linearly with V. This effect is seen by the points indicating traffic moment (VT) and the those indicating the predictions of Stott's hypothesis being close to each other on the graph.

The same effect can also be seen in Figure 5.2 which also relates to this class of level crossing.

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
	Critical values	0.1011055	0.10771987	0.11527916	0.12850792	0.15402052
Calculated values:		Implication for null hypothesis for Kolmogorov Smirnov test				
VT:	0.17151037	Reject	Reject	Reject	Reject	Reject
PD:	0.65088092	Reject	Reject	Reject	Reject	Reject
Stott:	0.19295071	Reject	Reject	Reject	Reject	Reject

5.2 Passive Vehicular Public - Network Rail accident data

Results from analysis of the 112 level crossings categorised as *passive vehicular public*.

Graphical results

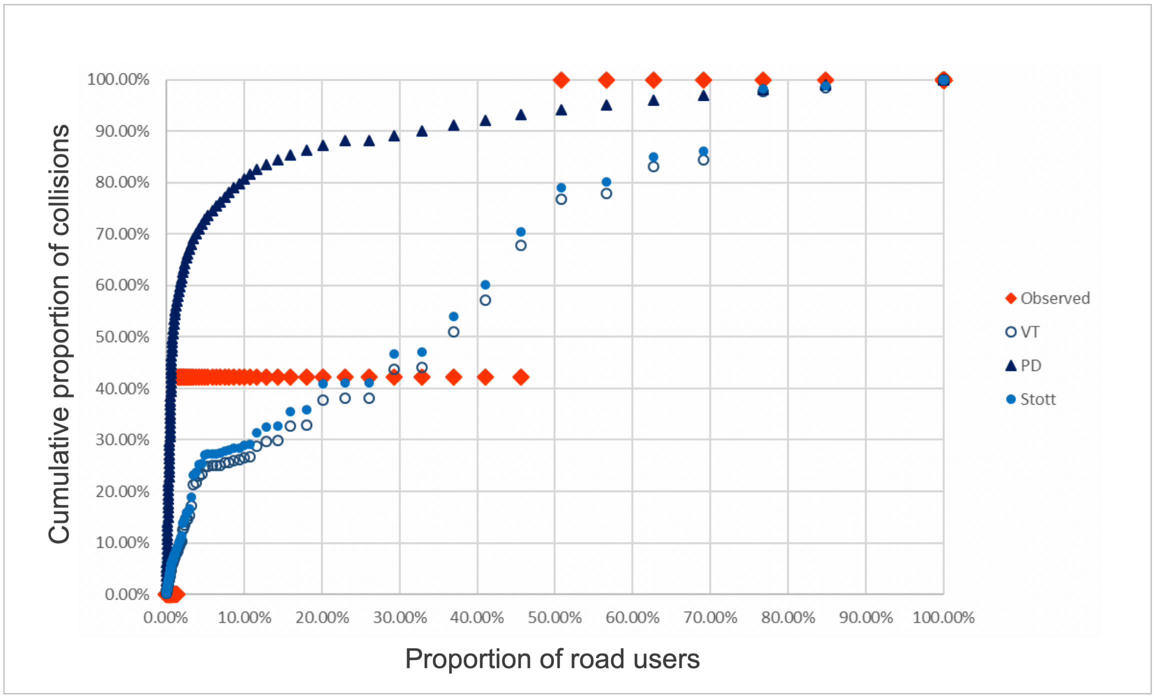


Figure 5.2: Observed accidents for passive vehicular level crossings using Network Rail data

For this class of level crossing, there are only two recorded collisions in the Network Rail data source. The effect of this small number of collisions is seen by the presences of two bands of orange square markers showing the observed collisions. One of the recorded collisions occurred at a level crossing with a small number of road users, which is seen by the lower band of orange square markers starting near to the left-hand edge of the graph.

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
Critical values		0.1011055	0.10771987	0.11527916	0.12850792	0.15402052
Calculated values:						
VT:	0.33776669	Reject	Reject	Reject	Reject	Reject
PD:	0.56946475	Reject	Reject	Reject	Reject	Reject
Stott:	0.32960806	Reject	Reject	Reject	Reject	Reject

5.3 Passive Vehicular Private or Staff - SMIS collision data

Results from analysis of the 2114 level crossings categorised as *passive vehicular private or staff.*

Graphical results

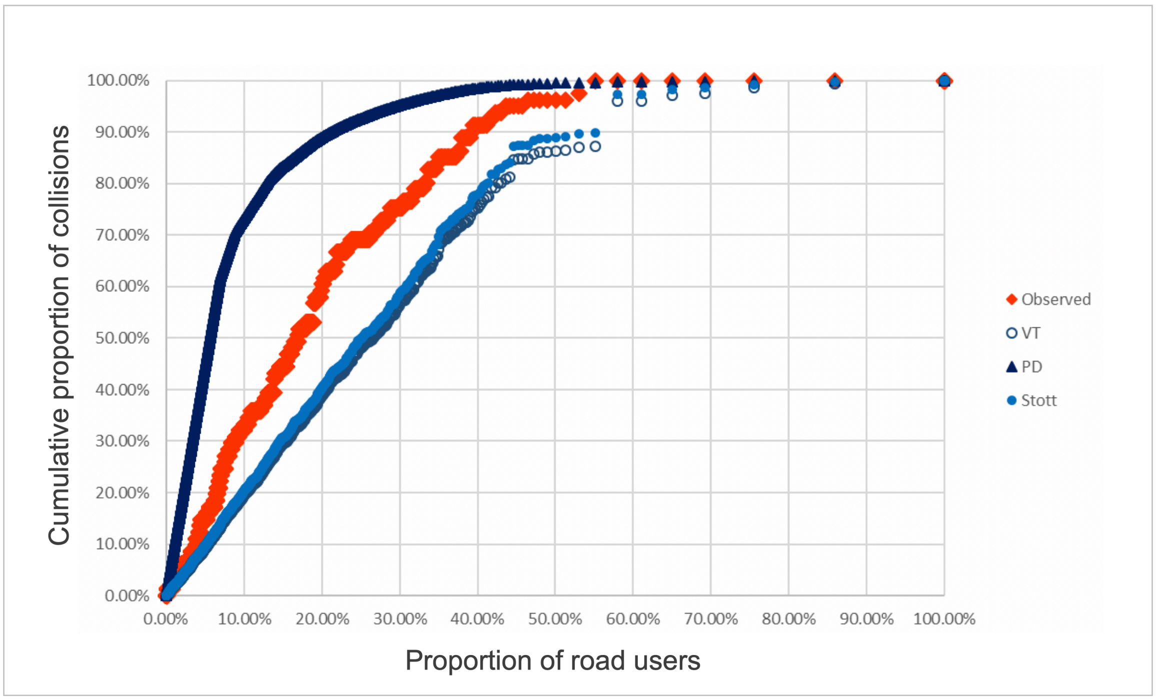


Figure 5.3: Observed collisions for passive vehicular private or staff level crossings using SMIS data

This class of level crossings occur only where there are few road users per day. Consequently a similar effect is seen in Figure 5.3 to the effect seen in Figures 5.1 and 5.2. For low values of *V*, Stott's hypothesis predicts that collision rise approximately linearly with V, consequently the points indicating the predictions of Stott's hypothesis and traffic moment (VT) are seen close to each other in Figure 5.3. The same effect is seen in Figure 5.4.

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
Critical values		0.02327187	0.02479433	0.02653428	0.0295792	0.03545154
Calculated values:						
VT:	0.24278548	Reject	Reject	Reject	Reject	Reject
PD:	0.42468071	Reject	Reject	Reject	Reject	Reject
Stott:	0.22695449	Reject	Reject	Reject	Reject	Reject

5.4 Passive Vehicular Private or Staff - Network Rail accident data

Results from analysis of the 2114 level crossings categorised as *passive vehicular private or staff*.

Graphical results

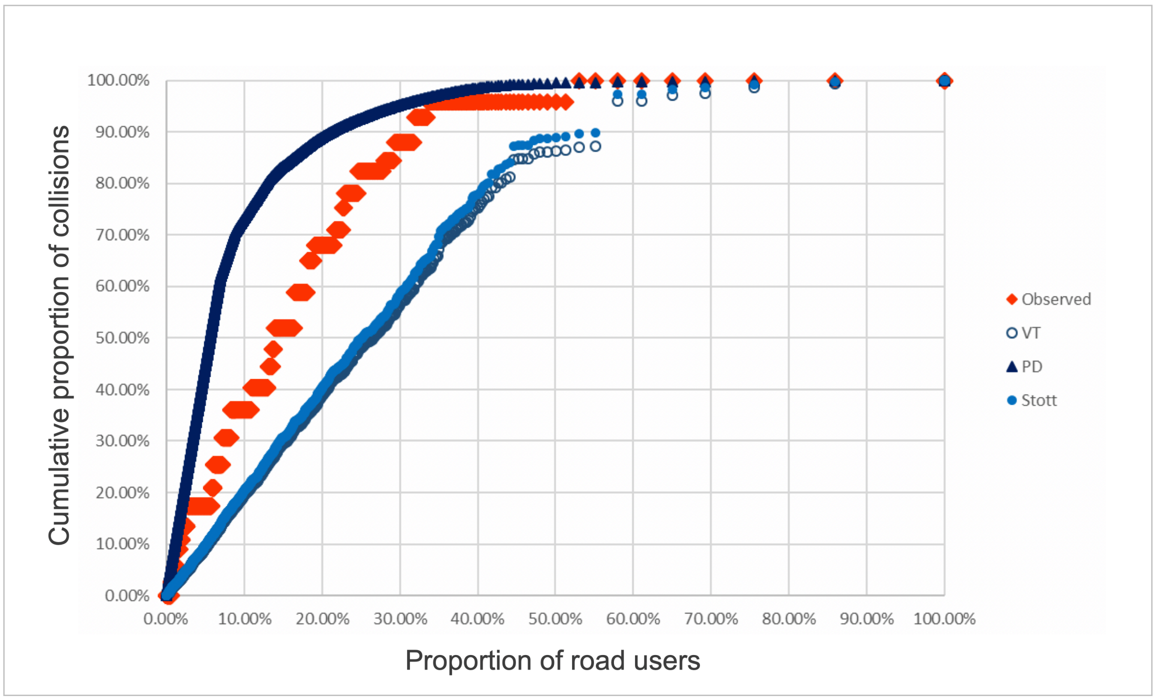


Figure 5.4: Observed accidents for passive vehicular private or staff level crossings using Network Rail data

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
Critical values		0.02327187	0.02479433	0.02653428	0.0295792	0.03545154
Calculated values:						
VT:	0.3463437	Reject	Reject	Reject	Reject	Reject
PD:	0.39988573	Reject	Reject	Reject	Reject	Reject
Stott:	0.32869292	Reject	Reject	Reject	Reject	Reject

5.5 Automatic Vehicular Public - SMIS collision data

Results from analysis of the 579 level crossings categorised as *automatic vehicular public.*

Graphical results

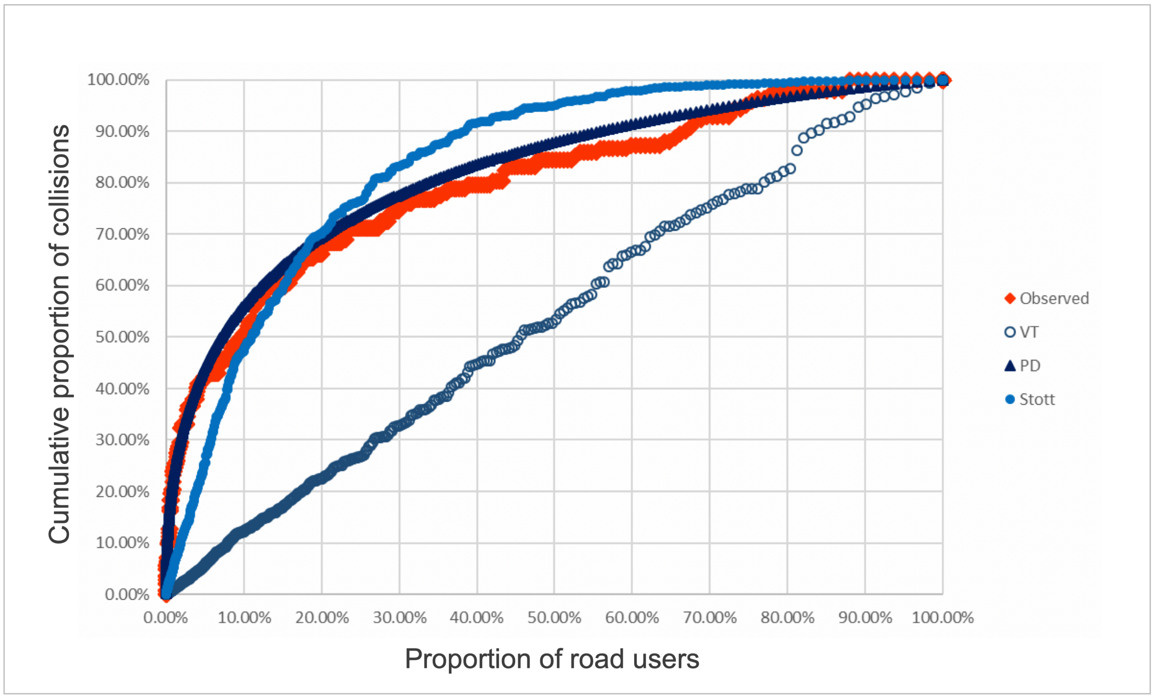


Figure 5.5: Observed collisions for automatic vehicular public level crossings using SMIS data

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
Critical values		0.04446768	0.04737678	0.05070147	0.05651967	0.06774049
Calculated values:						
VT:	0.45391305	Reject	Reject	Reject	Reject	Reject
PD:	0.06520384	Reject	Reject	Reject	Reject	**ACCEPT**
Stott:	0.22437973	Reject	Reject	Reject	Reject	Reject

5.6 Automatic Vehicular Public - Network Rail accident data

Results from analysis of the 579 level crossings categorised as *automatic vehicular public.*

Graphical results

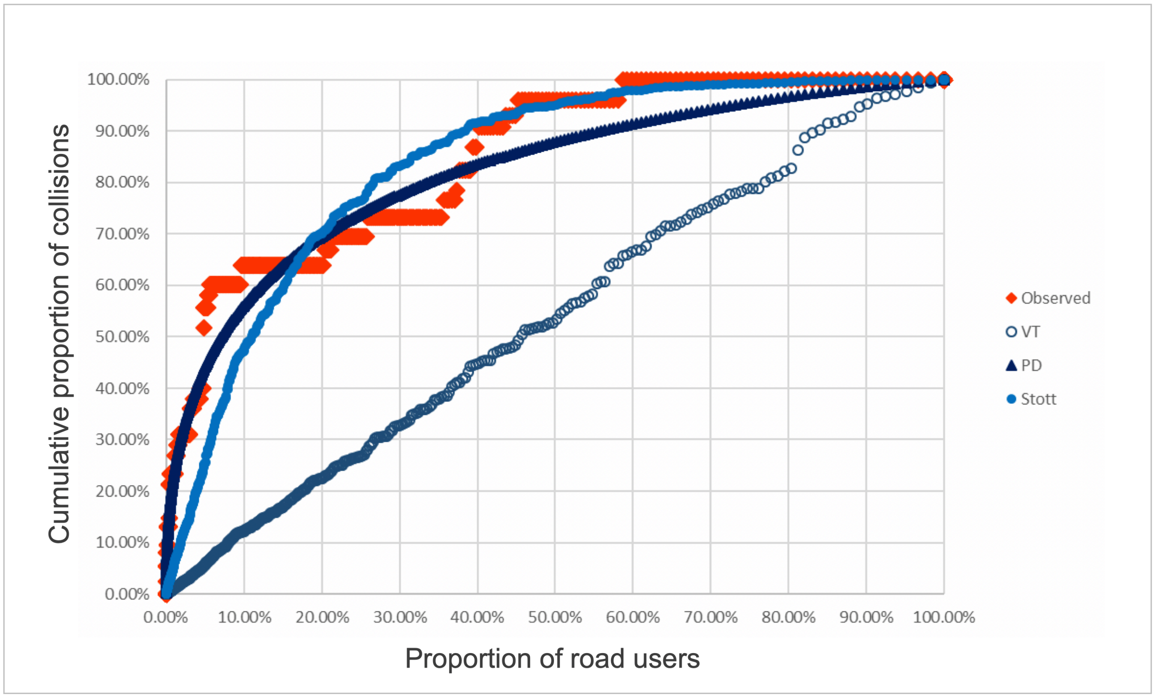


Figure 5.6: Observed accidents for automatic vehicular public level crossings using Network Rail data

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
Critical values		0.04446768	0.04737678	0.05070147	0.05651967	0.06774049
Calculated values:						
VT:	0.53426169	Reject	Reject	Reject	Reject	Reject
PD:	0.13992182	Reject	Reject	Reject	Reject	Reject
Stott:	0.30874185	Reject	Reject	Reject	Reject	Reject

5.7 Automatic Vehicular Private or Staff - SMIS collision data

Results from analysis of the 112 level crossings categorised as *automatic vehicular private or staff.*

Graphical results

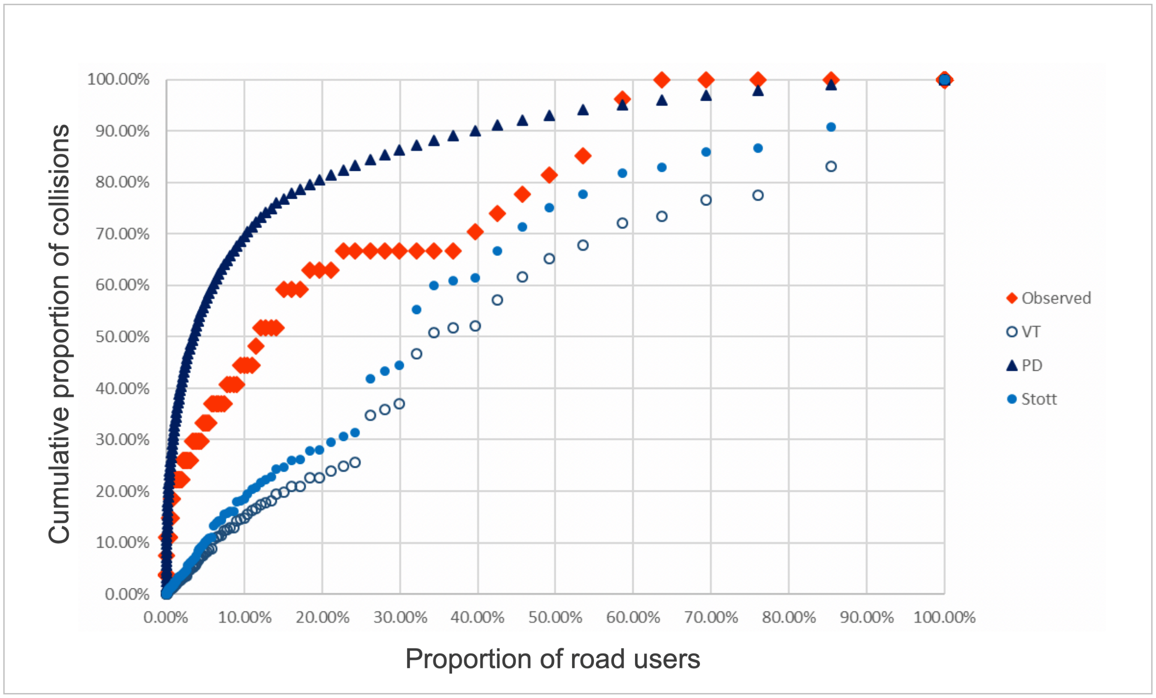


Figure 5.7: Observed collisions for automatic vehicular private or staff level crossings using SMIS data

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
Critical values		0.1011055	0.10771987	0.11527916	0.12850792	0.15402052
Calculated values:						
VT:	0.41796901	Reject	Reject	Reject	Reject	Reject
PD:	0.27003533	Reject	Reject	Reject	Reject	Reject
Stott:	0.36019325	Reject	Reject	Reject	Reject	Reject

5.8 Automatic Vehicular Private or Staff - Network Rail accident data

Results from analysis of the 112 level crossings categorised as *automatic vehicular private or staff.*

Graphical results

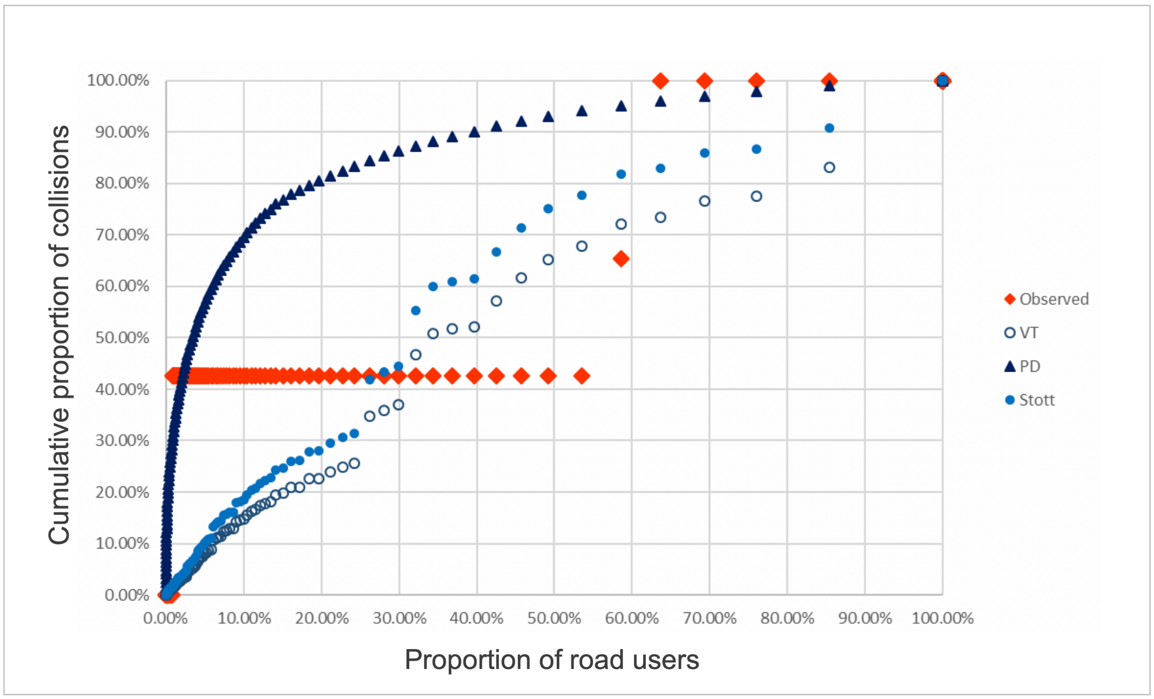


Figure 5.8: Observed accidents for automatic vehicular private or staff level crossings using Network Rail data

A similar effect is seen in Figure 5.8 to the effect seen in Figure 5.2; only three collisions are recorded in the Network Rail data for this class of level crossing which gives rise to the horizontal bands of orange square markers.

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
Critical values		0.1011055	0.10771987	0.11527916	0.12850792	0.15402052
Calculated values:						
VT:	0.41220943	Reject	Reject	Reject	Reject	Reject
PD:	0.51529099	Reject	Reject	Reject	Reject	Reject
Stott:	0.40858888	Reject	Reject	Reject	Reject	Reject

5.9 Railway-controlled Vehicular Public - SMIS collision data

Results from analysis of the 774 level crossings categorised as *railway-controlled vehicular public.*

Graphical results

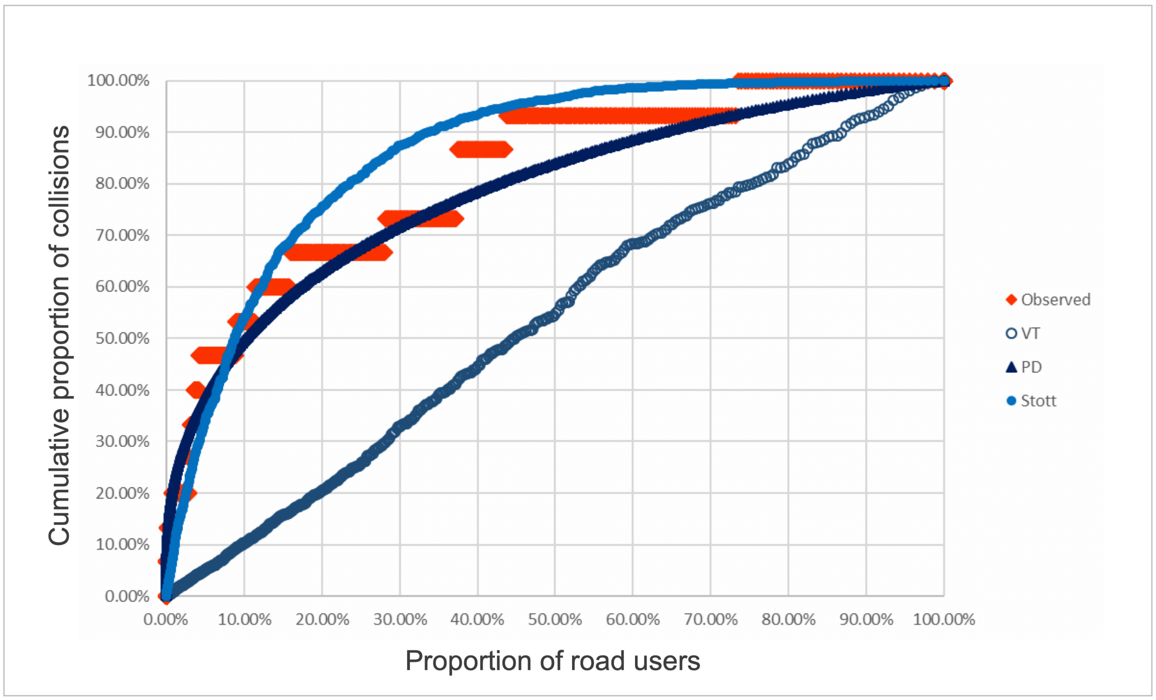


Figure 5.9: Observed collisions for railway-controlled vehicular public level crossings using SMIS data

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
Critical values		0.03846036	0.04097645	0.04385199	0.04888419	0.05858914
Calculated values:						
VT:	0.50275472	Reject	Reject	Reject	Reject	Reject
PD:	0.12546143	Reject	Reject	Reject	Reject	Reject
Stott:	0.18683914	Reject	Reject	Reject	Reject	Reject

5.10 Railway-controlled Vehicular Public - Network Rail accident data

Results from analysis of the 774 level crossings categorised as *railway-controlled vehicular public.*

Graphical results

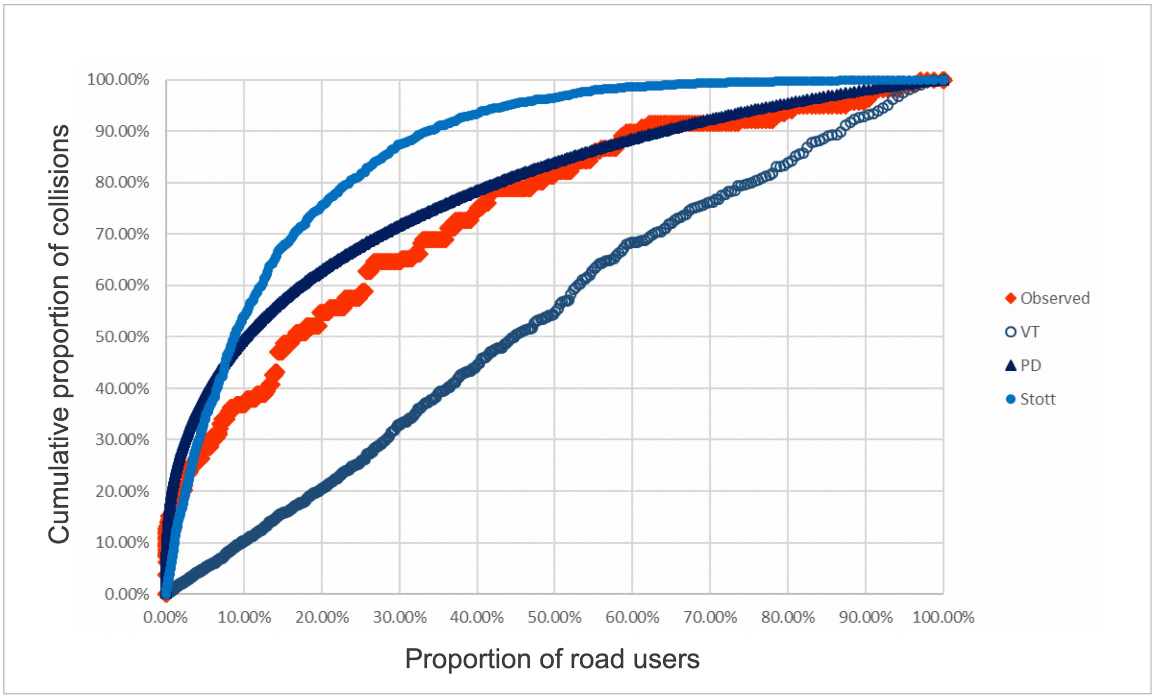


Figure 5.10: Observed accidents for railway-controlled vehicular public level crossings using Network Rail data

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
	Critical values	0.03846036	0.04097645	0.04385199	0.04888419	0.05858914
Calculated values:						
VT:	0.3633436	Reject	Reject	Reject	Reject	Reject
PD:	0.15183024	Reject	Reject	Reject	Reject	Reject
Stott:	0.23757215	Reject	Reject	Reject	Reject	Reject

5.11 Railway-controlled Vehicular Private or Staff - SMIS collision data

Results from analysis of the 51 level crossings categorised as *railway-controlled vehicular private or staff.*

Graphical results

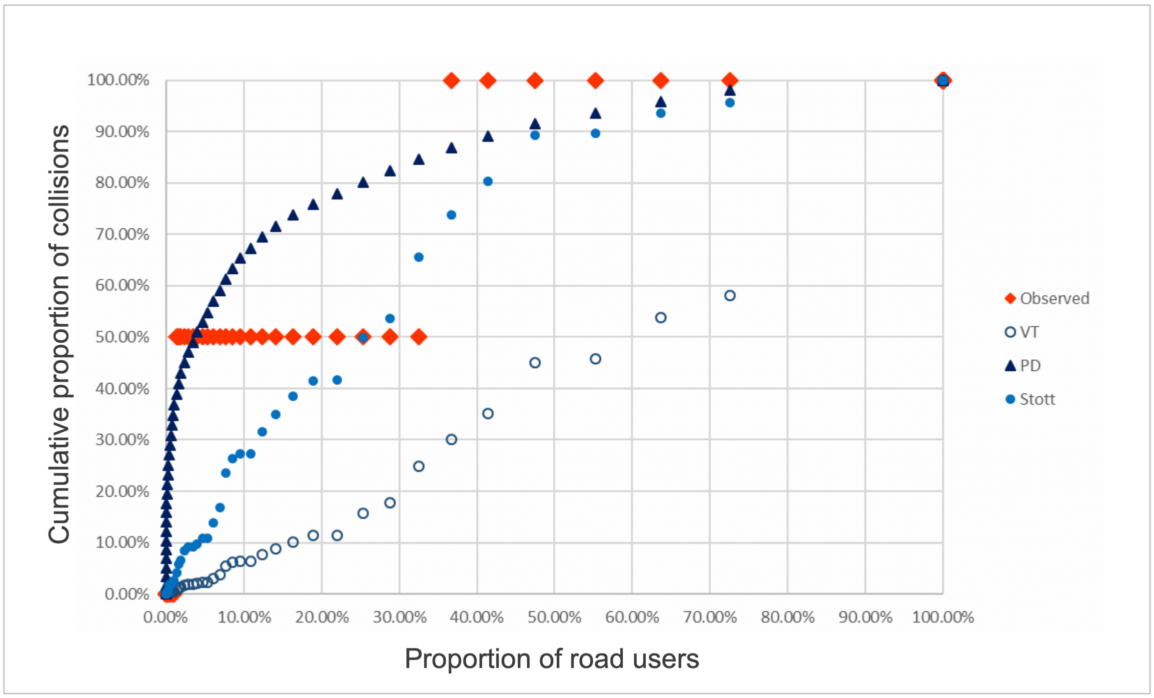


Figure 5.11: Observed collisions for railway-controlled vehicular private or staff level crossings using SMIS data

Figure 5.11 shows the same effect as noted for Figures 5.2 and 5.8. There are only two collisions recorded for this class of level crossing in the SMIS database, consequently horizontal bands of orange square markers are seen in the graph. Similarly there is only a single collision recorded in the Network Rail data source for this class of level crossing giving rise to the same horizontal banding being seen in Figure 5.8.

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
Critical values		0.14982997	0.15963193	0.17083417	0.19043809	0.22824565
Calculated values:						
VT:	0.69890993	Reject	Reject	Reject	Reject	Reject
PD:	0.36744083	Reject	Reject	Reject	Reject	Reject
Stott:	0.45870528	Reject	Reject	Reject	Reject	Reject

5.12 Railway-controlled Vehicular Private or Staff - Network Rail data

Results from analysis of the 51 level crossings categorised as *railway-controlled vehicular private or staff.*

Graphical results

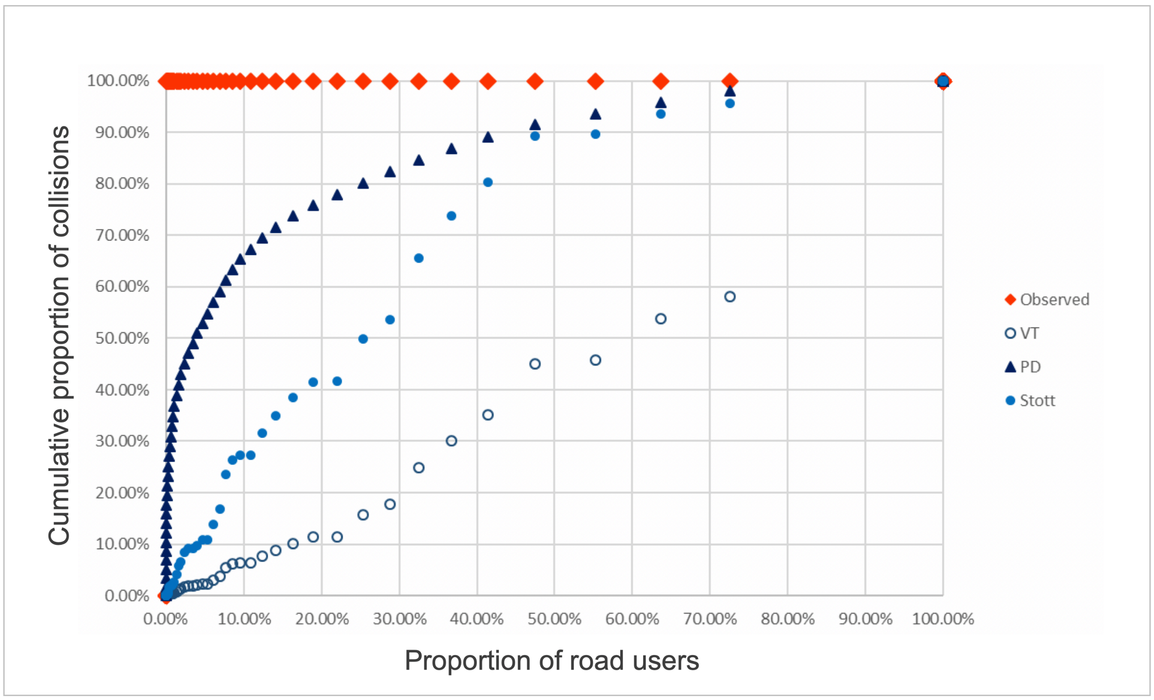


Figure 5.12: Observed accidents for railway-controlled vehicular private or staff level crossings using Network Rail data

Tabulated results

		Alpha value				
		20%	**15%**	**10%**	**5%**	**1%**
Critical values		0.14982997	0.15963193	0.17083417	0.19043809	0.22824565
Calculated values:						
VT:	0.99954862	Reject	Reject	Reject	Reject	Reject
PD:	0.8244836	Reject	Reject	Reject	Reject	Reject
Stott:	0.99768152	Reject	Reject	Reject	Reject	Reject

5.13 Differences over ranges of traffic volumes

It is noteworthy that the curves for the cumulative numbers of predicted collisions appear to change shape in the various tests. For example in Figures 5.1 to 5.4 the curves for the number of collisions predicted by Stott's hypothesis lies close to the traffic moment curve, whereas in other graphs, these curves are not closely aligned. This effect arises because of the differing range of road vehicles considered in the various tests. Table 5.2 shows the maximum count of vehicular road users per day for each category of level crossing.

Table 5.2: Maximum count of vehicular road users per day for each category of level crossing

Level crossing category	Maximum count of vehicular road users per day
Passive vehicular public	570
Automatic Vehicular Private or Staff	1,253
Passive Vehicular Private or Staff	1,350
Railway-controlled Vehicular Private or Staff	12,069
Automatic vehicular public	17,550
Railway-controlled Vehicular Public	29,592

Whilst the traffic moment model predicts that collisions should increase linearly over the range of vehicular road traffic (V), the Stott and Peabody Dimmick models predict non-linear collision rates over V. Figure 2.6 shows that for low values of V the predicted collision rate increases nearly linearly. The graph for passive vehicular level crossings has a low range of values of V (from 0 to 570 as shown in Table 5.2), over this low range of values the collision predictions of the Stott model is nearly linear. This effect can be seen in Figures 5.1 and 5.2 which shows that the Stott model and the traffic moment model align closely with each other. Conversely the range of V for railway-controlled vehicular public level crossings is much wider (from zero to 29,592, as shown in Table 5.2). Over this wide range the collision rate predicted by Stott's model increases then decreases (Figure 2.6). For these level crossings, shown in Figures 5.9 and 5.10, it can be seen that the curves for traffic moment and Stott's model are not closely aligned.

5.14 Roughness of the traffic moment curve

In each of the graphs in Figures 5.1 to 5.12 the curve for traffic moment is approximately, although not exactly, a straight line. The lack of smoothness arises since traffic moment is the product of road vehicles and trains traversing a level crossing, whereas the independent variable on the graph's abscissa is only the number of traverses by road vehicles (V). Since the number of trains per day will vary between level crossings, from point to point along the abscissa, the line will move in an unpredictable way, which is what is seen in the graphs. It can be expected that there is some correlation between road traffic volumes and train volumes at a level crossing: in busy urban areas it is likely that level crossings will have a high volume of both road vehicles and trains. Conversely in rural areas it may be expected that there would be few vehicles of either mode. However the correspondence will not be exact, for example a private level crossing in rural area over a busy mainline rail could have few road vehicles yet many trains. In general there is no reason to expect that traffic moment will increase linearly with increasing road traffic volumes. Overall the cumulative measure of traffic moment will always increase with increasing traffic volumes, however the exact increase at any point on the graph will depend on train volumes at each level crossing, which may be independent of road traffic volumes. The effect is a non-linear increase in cumulative traffic moment over the range of traffic volumes.

5.15 Application of the test where there are few recorded collisions

For some classes of level crossings, there are cases where there is small number of recorded collisions. For example Figure 5.8 shows only three recorded collisions, which can be seen as a step shape in the orange square markers; a similar effect can also be seen in Figures 5.11 and 5.12. These graphs show the results for *automatic* and *railway-controlled* level crossings on private land or that are used only by railway staff. The majority of private or staff level crossings are passive, and therefore there are relatively few level crossings in these classes. Furthermore automatic and railway-controlled level crossings have relatively lower safety risk than passive level crossings, which is the reason that there are few recorded collisions for level crossings in these classes.

Whilst the observed collisions in these cases show a stepped shape, the traffic models all describe a smooth variation in safety risk over the range of traffic values. Consequently it cannot be expected that there will be a good correlation between the observation and the traffic models, which is what is seen in the results of the Kolmogorov Smirnov test for these cases. Such a lack of correlation is a natural consequence of applying statistical methods to data where there are only a small number of observations: in general, it is meaningful to apply statistical methods only in cases where there is a large number of samples. As discussed in Section 1.4, all classes of level crossing have been included in this study for the sake of completeness, even though the results may not have good statistical value in cases where there are few recorded collisions

5.16 Summary of results

Table 5.3 provides a summary of the results of the Kolmogorov Smirnov test for each of the traffic models for the 12 cases tested. In the table, lower values of the Kolmogorov Smirnov test result indicate a better correlation between the data sets. The lowest result is 0.0652 for the Peabody Dimmick traffic model for automatic vehicular public level crossing using SMIS collision data; this is the only test that shows a statistically significant correlation and only with a value of $\alpha = 1\%$. There is a large range in the results with the largest results being more than 0.99, however these are for a category that contains only 51 level crossings and a small number of recorded collisions and therefore may have too few data points for a meaningful analysis to be performed.

Table 5.3: Summary of the results of the Kolmogorov Smirnov test for each of the traffic models

Test	Category	Number of level crossings	Max number of road vehicles	Data source	Kolmogorov Smirnov test			
					Results			Threshold at α = 1%
					Peabody Dimmick	Stott	Traffic moment	
1	Passive vehicular public	112	570	SMIS	0.6509	0.1930	0.1715	0.1540
2				Network Rail	0.5695	0.3296	0.3378	
3	Passive vehicular private or staff	2114	1350	SMIS	0.4247	0.2270	0.2428	0.0355
4				Network Rail	0.3999	0.3287	0.3463	
5	Automatic vehicular public	579	17,550	SMIS	**0.0652**	0.2244	0.4539	0.0677
6				Network Rail	0.1399	0.3087	0.5343	
7	Automatic vehicular private or staff	112	1253	SMIS	0.2700	0.3602	0.4180	0.1540
8				Network Rail	0.5153	0.4086	0.4122	
9	Railway-controlled vehicular public	774	29,592	SMIS	0.1255	0.1868	0.5028	0.0586
10				Network Rail	0.1518	0.2376	0.3633	
11	Railway-controlled vehicular private or staff	51	12,069	SMIS	0.3674	0.4587	0.6989	0.2282
12				Network Rail	0.8245	0.9977	0.9995	

5.17 Overall correlation of the traffic models

An immediate interpretation of the results may be that none of the traffic models correlates well with observed collision nor accidents, but that the Peabody Dimmick model performs best amongst the set of poor candidates. Such an interpretation would, however, be simplistic and misleading. Figure 5.13 shows the range of results from the Kolmogorov Smirnov tests grouped by traffic model. The figure shows that each of the models has a results of less than 0.2 for at least one of the tests; similarly each model has a test result of more than 0.8 for at least one test. As such it can be seen that in some cases each of the tests correlates reasonably well with the observation data and in other tests it does not. Overall it could be concluded that none of the models performs significantly better than any other.

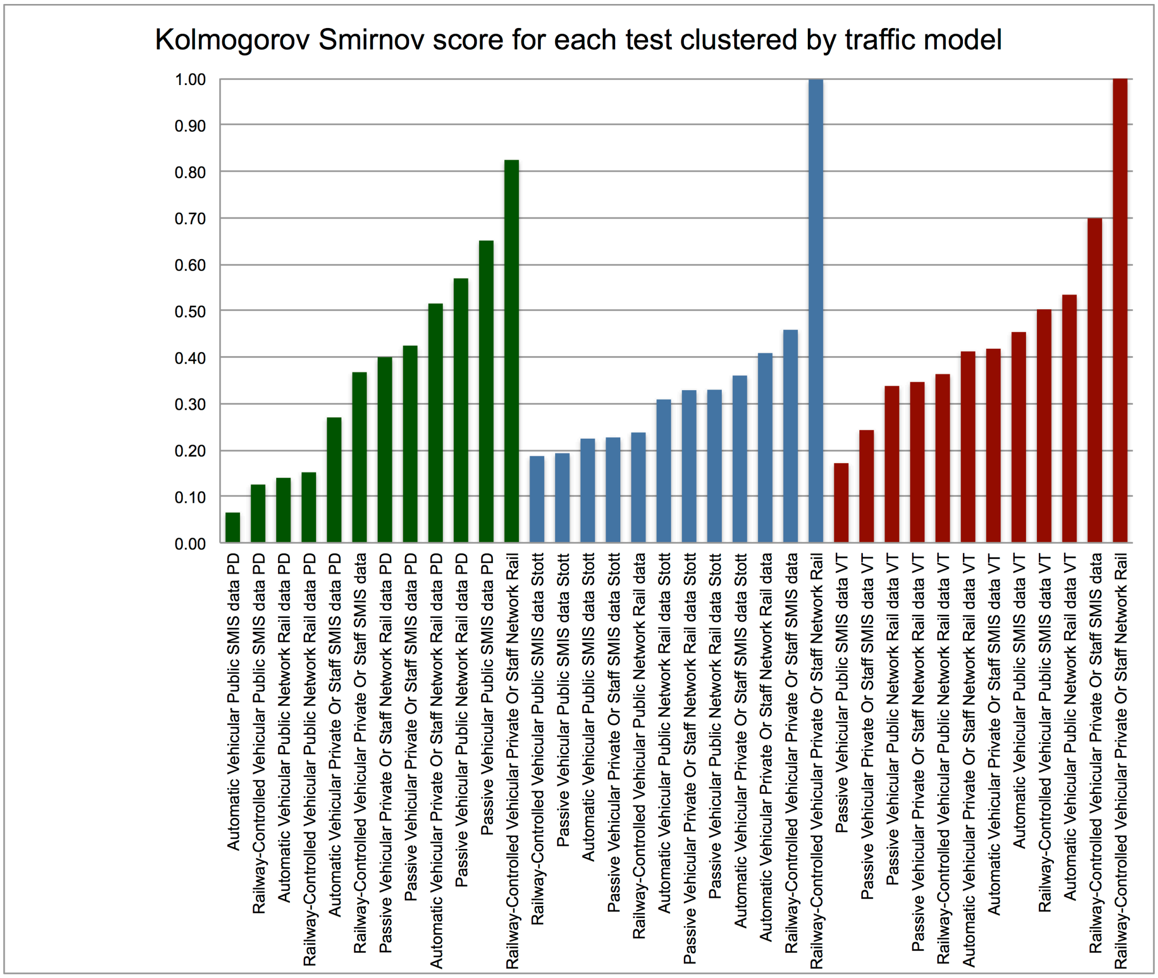


Figure 5.13: Range of Kolmogorov Smirnov results by traffic model

5.18 Statistical power by number of level crossings

As previously noted, collisions at level crossings are – fortunately – relatively rare events. In general statistical tests are more powerful when there are more data available. It may therefore be expected that where there are more level crossings in the test sample then it would be possible to obtain better correlation with a traffic model. Figure 5.14 shows a scatter plot of Kolmogorov Smirnov test values over the number of level crossings in the test sample. It can be seen that the largest Kolmogorov Smirnov scores are obtained in the categories that have fewest level crossings which may indicate that when there are fewer data points there is less opportunity for correlation to be found in the data. However the converse effect is not seen: the tests with the largest number of level crossings do not have the lowest test values. In fact the lowest result is obtained in the category that has only 579 level crossings, significantly lower than the 2114 level crossings in the most populous category. Overall it can be seen that poor correlation of the traffic models against observed collisions cannot generally be attributed to the small number of level crossings in the test categories.

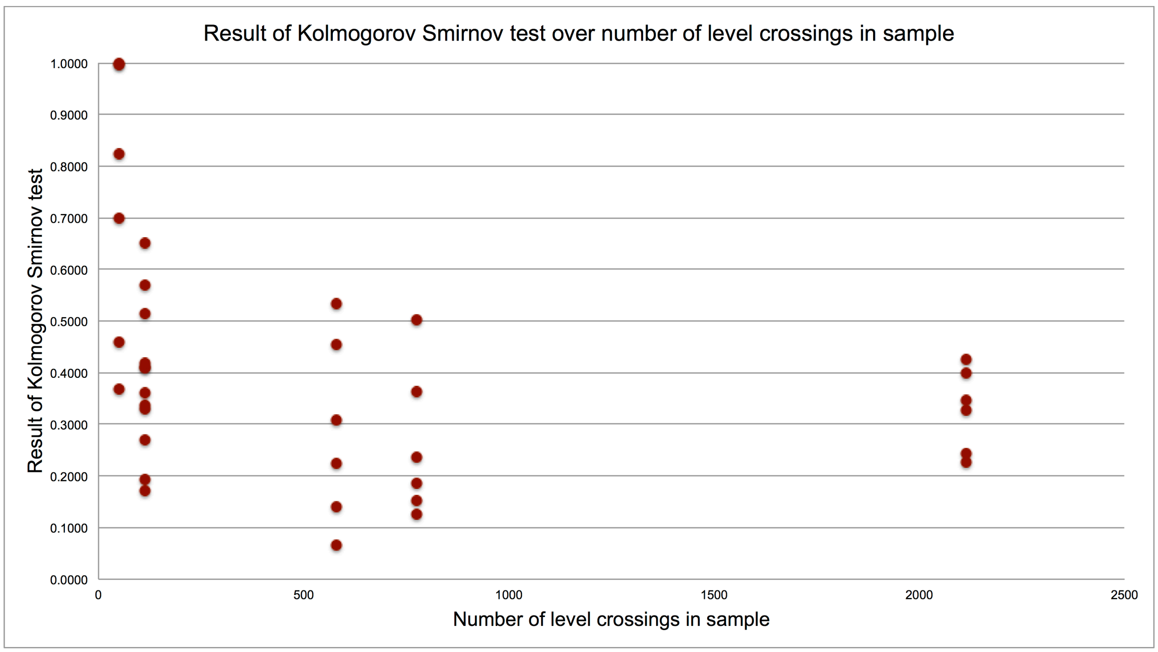


Figure 5.14: Scattering of Kolmogorov Smirnov results compared with number of level crossings in the test sample

5.19 Correlation of models by road traffic volume

As noted above, for low values of road traffic volume (V), the Stott hypothesis proposes that collisions should increase with road traffic volumes in a manner similar to the increase with traffic moment. For higher values of V, there is a greater difference between the models. It therefore needs to be considered whether the correlation between the traffic models and observed collisions varies by the maximum value of V in the test. Specifically there is a question whether one model consistently performs better for lower values of V and another consistently performs better for high values of V.

Looking at the results it can be seen, for example, in Figure 5.1 where there are relatively low values of V, there appears to be a good correlation between observed collisions and the traffic moment and Stott models. Conversely in Figure 5.5 where there are larger values of V, there is the best correlation with the Peabody Dimmick model.

Table 5.4 shows the six categories of level crossing ranked by ascending value of maximum V together with the traffic model which had the lowest test value using either the SMIS collision data or Network Rail accident data. The table shows that overall the Peabody Dimmick model provides the best correlation although, as noted above, there is a range of test values and overall there is not good correlation. The test with the lowest maximum value of V correlated best with traffic moment, and the test with the highest maximum value correlated best with the Peabody Dimmick model. Overall the results are inconclusive, it does not appear that there is a clear pattern that correlation is a function of V.

Table 5.4: Best correlating model by level crossing category ranked by maximum *V* using SMIS data

Category	Maximum road traffic volume (*V*)	Best correlating model	Kolmogorov Smirnov test	
			Result	Threshold at α = 1%
Passive Vehicular Public	570	Traffic moment	0.1715	0.1540
Automatic Vehicular Private or Staff	1253	Peabody Dimmick	0.2700	0.1540
Passive Vehicular Private or Staff	1350	Stott	0.2270	0.0355
Railway-controlled Vehicular Private or Staff	12,069	Peabody Dimmick	0.3674	0.2282
Automatic Vehicular Public	17,550	Peabody Dimmick	0.0652	0.0677
Railway-controlled Vehicular Public	29,592	Peabody Dimmick	0.1255	0. 0586

5.20 Implications of poor correlation for SRPTs

This study has considered the correlation between observed collisions and the rates predicted by *traffic models*, notably the study has not been able to test the correlation between observed collisions and SRPTs since, in most cases, there are no details on the methods of calculation in the SRPTs. It is theoretically possible that the poor correlation between collision rates and traffic models does not extend to a poor correlation with the SRPTs. The traffic models are only one aspect of the tools and, in most cases, the tools consider the effects of other factors on the overall risk predictions. For example, it is possible that the GB railways' ALCRM model provides very accurate predictions despite using Stott's hypothesis as the underlying traffic model: errors between actual collision rates and the rate predicted by Stott may be corrected by other (unspecified) calculations in the tool.

It is noted that the ALCRM initially used traffic moment as its underlying traffic model, but a change was made to adopt Stott's hypothesis instead (Baker and Heavisides, 2007). No evidence is available to say whether the change to traffic model increases the predictive accuracy of the ALCRM, however it must be presumed that the change was motivated by the desire to improve the accuracy of the tool. It therefore appears that the amount of correction that is possible with an SRPT is limited and the choice of underlying traffic model has an impact on the overall accuracy of the tool.

Since the Peabody Dimmick model appears overall to be better correlated with collision rates at British level crossings, it could be argued that better results might be achieved if it had been selected instead of Stott's hypothesis. However it must be considered that the degree of correlation between the Peabody Dimmick model and observed collisions is not especially good. Furthermore Stott's hypothesis and Peabody Dimmick's model are only two traffic models from a hypothetical infinity of potential traffic models that could be used. It seems unnecessary to debate whether one poorly correlating traffic model should be selected in preference to any other poorly correlating model. A more robust approach would be to identify an appropriate model that has good correlation with observed collision. The review of the literature has identified only three models that can be tested, and the study described in this chapter has identified that none of these three models are good candidates for consideration. This point is considered further in the next chapter.

5.21 Contribution

The following contributions to current knowledge have been made in this chapter:

Contribution 6: The study described in this chapter has undertaken rigorous tests to compare observed collisions with proposed traffic models. The method was performed in a repeatable manner that would allow the test to be carried out with any other traffic model that may be proposed.

Contribution 7: Further to the finding in Chapter 3 which showed that traffic moment can be a normaliser for collisions in idealised conditions, in real-world conditions it does not appear that observed collisions vary in accordance with traffic moment. This finding may have profound implications for the many SRPTs that use traffic moment as an underlying traffic model.

Contribution 8: Similarly, it does not appear that in general Stott's hypothesis is a meaningful normaliser for observed traffic collisions. Specifically Stott's hypothesis was developed to describe collisions at automatic vehicular level crossings, however the Peabody Dimmick model appears to correlate better with observation in this class of level crossing.

Contribution 9: The descriptive traffic model developed by Peabody Dimmick is no worse at describing the rate of level crossing collisions than the predictive hypothesis developed by Stott. This finding is particularly noteworthy since the Peabody Dimmick model was developed approximately 90 years ago. Again this finding may have profound implications for the SRPT that uses this traffic model.

Contribution 10: In general, none of the proposed traffic models correlated with observed collisions with any meaningful degree of statistical significance.

Chapter 6: Study of collision rates over number of road users

The approach of cumulating the values in the Kolmogorov Smirnov test provides an effective method to analyse overdispsersed data. However the use of cumulative values and normalising the values in a range between zero and one leads to a loss of information since the test does not compare the absolute values of the test variable and it is not possible to infer the absolute values from the normalised data. This section describes an analysis where collision data cumulated over road traffic volume is used, as in the Kolmogorov Smirnov test, however absolute values are used.

6.1 Test with absolute values

It is hypothetically possible that the absolute values in the test variable and the comparison distribution can vary by orders of magnitude, however as long as the distributions are similar in shape, the Kolmogorov Smirnov test can still show a high degree of correlation between the distributions. The accumulation and normalisation of the value also occurs over the range of the independent variable. In the correlation test described in Chapter 5 the range of traffic volumes are all scaled from zero to one in the tests. This scaling leads to the effect where the comparison distribution appears to change shape between the different tests due to the tests being conducted for a different range of traffic volumes (being the independent variable in the tests). In principle, the set of cumulative values used in the Kolmogorov Smirnov test could provide a useful data set that could be used for further analysis of level crossing collisions, however by scaling the values of the independent variable (road user volume) to be between zero and one, some information is lost that hinders further analysis.

Figures 6.1 to 6.6 show the cumulative rate of collisions per train traverse per day (being the unit determined in Section 4.4) over the absolute number of road users. These graphs are similar to the graphs shown in Figures 5.1 to 5.12, however the road traffic volumes have not been accumulated, and the values on neither scale have been normalised. The graphs shown in Figures 6.1 to 6.6 have a general shape that approximates a power function over traffic volume of the form of: $y = ax^b$, where a and b are constants. There is no underlying theory to suggest why a curve of this form should correlate with observed collisions, rather this equation has been selected for further exploration based on observation of the data.

Mathematical analysis software was used to perform a regression analysis determine a best fit power function to each curve. The results of the regression analysis are also shown on the figures. Table 6.1 shows the equation of the power curve determined during the regression analysis and the R^2 value of the fit between the curve and the observation data. Discussion of the results shown in Figures 6.1 to 6.6 is provided in Section 6.2.

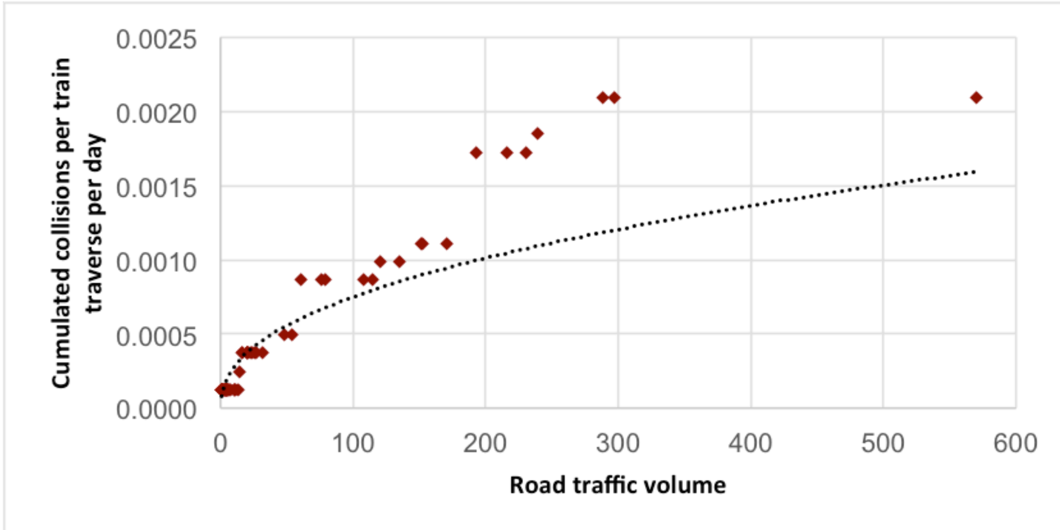


Figure 6.1: Cumulative collisions per train traverse per day over road traffic volume for Passive Vehicular Public level crossings

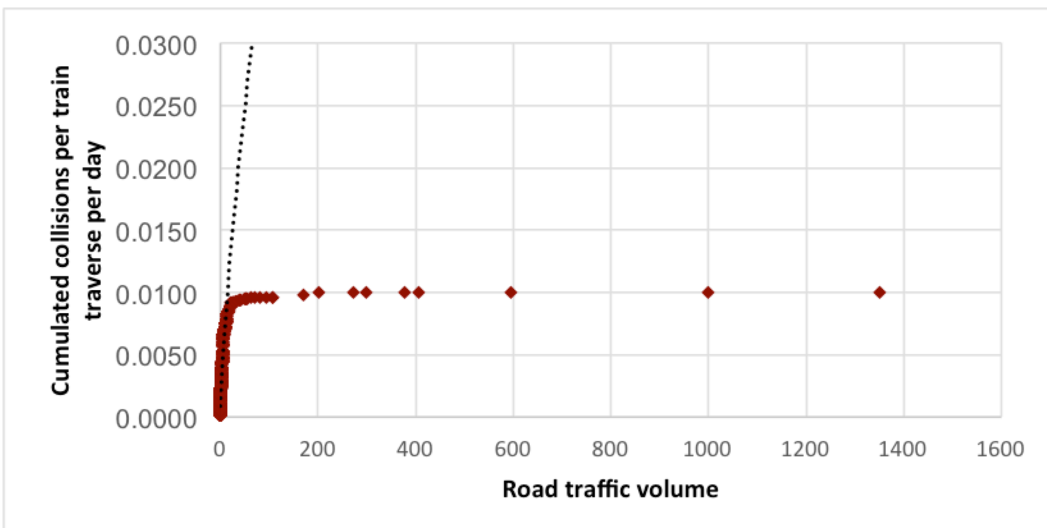


Figure 6.2: Cumulative collisions per train traverse per day over road traffic volume for Vehicular Private or Staff level crossings

In Figure 6.2, there is a concentration of level crossings with low values of road user volume: most of the points occur towards the left-hand edge of the graph, there are few points occurring at higher values of V. The best-fit line has been heavily influenced by these points at lower values creating an effect where for higher values of V there is a poor correspondence between the cumulative rates of observed collisions and the best-fit line. The same effect is seen, although to a lesser degree, in Figure 6.3.

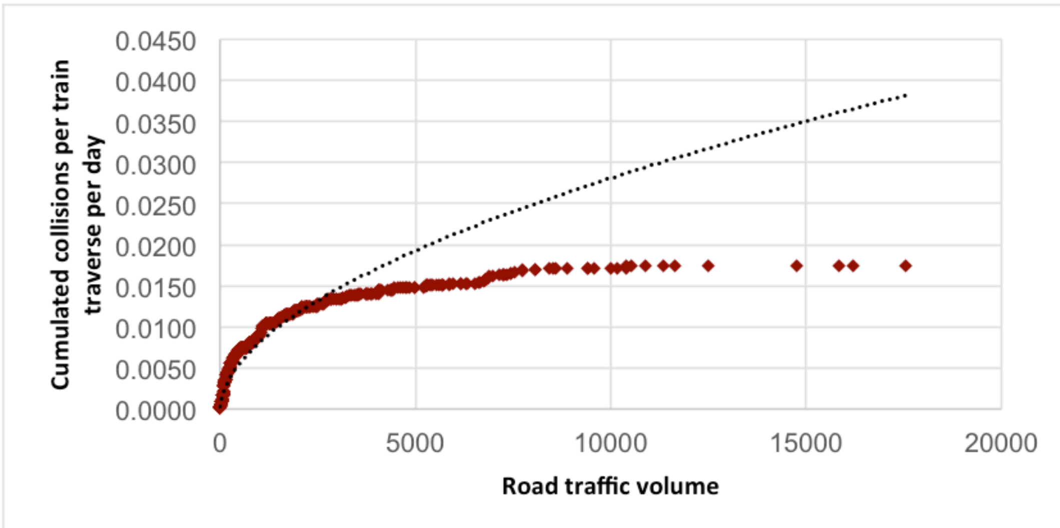


Figure 6.3: Cumulative collisions per train traverse per day over road traffic volume for Automatic Vehicular Public level crossings

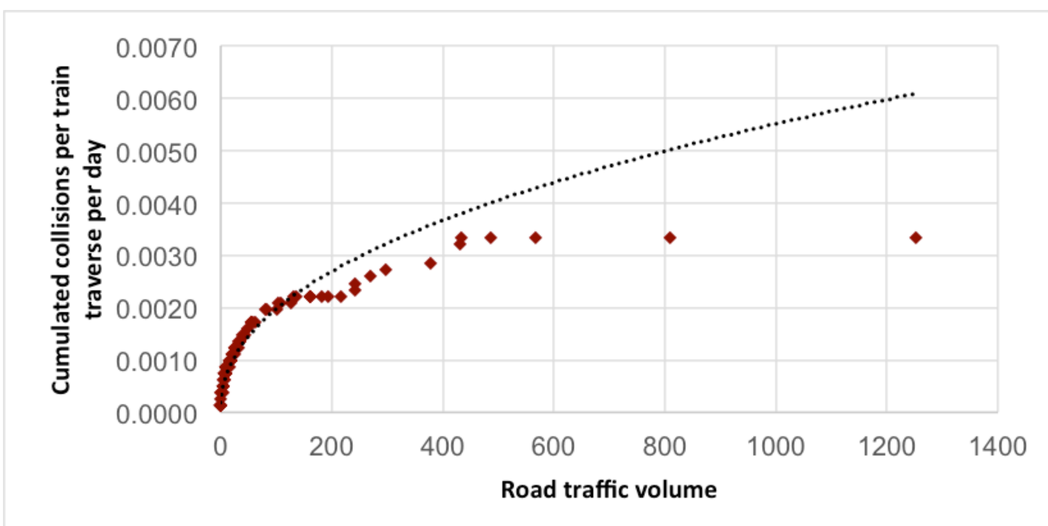


Figure 6.4: Cumulative collisions per train traverse per day over road traffic volume for Automatic Vehicular Private or Staff level crossings

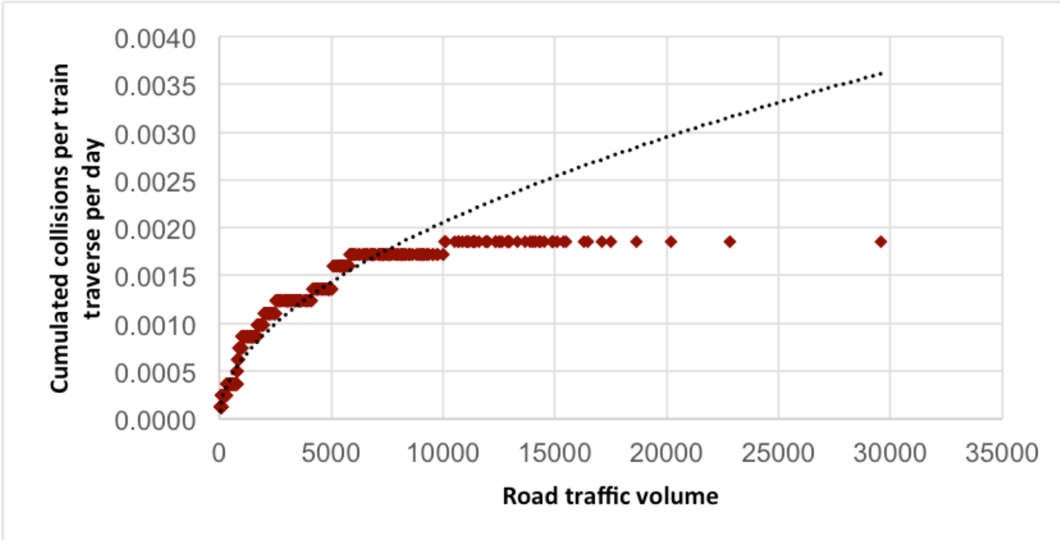


Figure 6.5: Cumulative collisions per train traverse per day over road traffic volume for Railway-controlled Vehicular Public level crossings

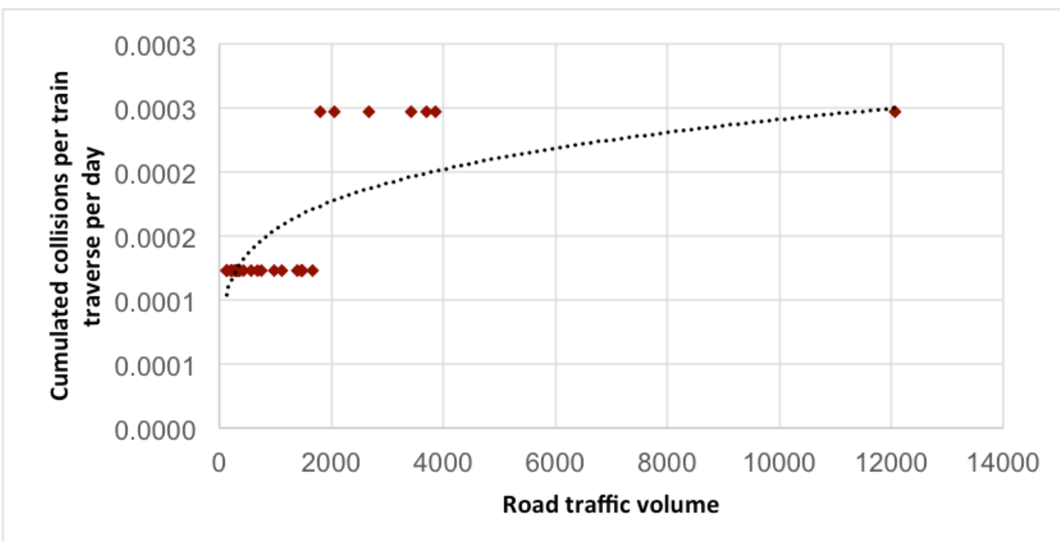


Figure 6.6: Cumulative collisions per train traverse per day over road traffic volume for Railway-controlled Vehicular Private or Staff level crossings

For this class of level crossing, there are only two collisions recorded in the SMIS database. Consequently the same effect is seen in Figure 6.6 as is seen in Figure 5.11 where there are horizontal bands of points indicating the number of observed collisions.

Table 6.1: Equations and R^2 values of the curves determined from the regression analysis

Graph	Equation of curve determined by regression analysis	R^2 value of regression curve fit to data
Figure 6.1: Passive Vehicular Public	$y = 1.0 \times 10^{-4} V^{0.432}$	*0.821*
Figure 6.2: Passive Vehicular Private or Staff	$y = 1.5 \times 10^{-3} V^{0.725}$	*0.551*
Figure 6.3: Automatic Vehicular Public	$y = 2.0 \times 10^{-4} V^{0.544}$	*0.919*
Figure 6.4: Automatic Vehicular Private or Staff	$y = 3.0 \times 10^{-4} V^{0.444}$	*0.944*
Figure 6.5: Railway-controlled Vehicular Public	$y = 2.0 \times 10^{-5} V^{0.521}$	*0.952*
Figure 6.6: Railway-controlled Vehicular Private or Staff	$y = 4.0 \times 10^{-5} V^{0.192}$	*0.572*

For smooth discrete functions, the cumulated value of the test variable is the *integral* of the function. Where the equation of the cumulated curve is known, then it is possible the determine the shape of the underlying distribution by differentiating cumulated function. With the values shown in Table 6.1, differentiating the equations over V would give an equation of the form:

$$CP(\text{collision per train traverse per day}) = a'V^{b'}$$

where:

CP	is the cumulative probability;
V	is the number of road users per day;
a' and b'	are constants determined during the differentiation.

Such an equation is an estimate of the rate at which collisions per train traverse per day vary over road user volumes for each category of level crossing. Table 6.2 shows the formulae that are determined from differentiating the results shown in Table 6.1.

Table 6.2: Differentiated results of regression curves described in Table 6.1

Level crossing category	***CP(collision per train traverse per day)*** **estimate determined from differentiation of formulae in Table 6.1**
Passive Vehicular Public	$4.3\times10^{-5}V^{-0.568}$
Passive Vehicular Private or Staff	$1.1\times10^{-3}V^{-0.275}$
Automatic Vehicular Public	$1.1\times10^{-4}V^{-0.456}$
Automatic Vehicular Private or Staff	$1.3\times10^{-4}V^{-0.556}$
Railway-controlled Vehicular Public	$1.0\times10^{-5}V^{-0.479}$
Railway-controlled Vehicular Private or Staff	$7.7\times10^{-6}V^{-0.808}$

6.2 Interpretation of non-zero collision predictions

As noted above, there are many level crossings where there is no history of collisions having occurred. It is this phenomenon that gives rise to the overdispersion seen in the data set. Conversely all traffic models predict non-zero collision rates for all level crossings. This apparent anomaly between observation and risk prediction is a common feature of safety risk management, especially for rare events. It can be seen in Table 6.2 that the predicted number of collisions for each level crossing is very small, for example the risk prediction in the first row of the table is $4.3\times10^{-5}V^{-0.568}$ per day, which means that even for a level crossing with a low value of V, the predicted number of collisions will never be less than 4.3×10^{-5}, furthermore as V increases the predicted rate reduced further. Clearly it is not possible to have 10^{-5} collisions per day: collisions are discrete events that can occur in only whole numbers. The predicted small numbers of collisions means that there will be many level crossings where there remains a low risk of a collision occurring, even though no collisions have yet been observed. In order to validate the correctness of risk predictions for rare events it is necessary to aggregate the observations over either a large period of time at individual level crossings, or across a large sample of level crossings over a small period.

6.3 Collection of near-collision data

Overfitting is an effect that can occur during regression analysis when a curve has been fitted to data in such a way as to make the residual error between the model and the observations smaller than the natural variation in the data. An overfitted model describes noise in the sample rather than the general trend, and can have poor predictive accuracy for unseen data samples. Figure 6.7 shows an example of overfitting. The relationship shown by the data in the figure is a simple power function where some random error has been introduced to the data. Figure 6.7a shows the actual relationship between the data when the random variation has been removed; Figure 6.7b shows an example of an overfitted curve that has better correlation with the source data, but incorrectly describes the noise in the data.

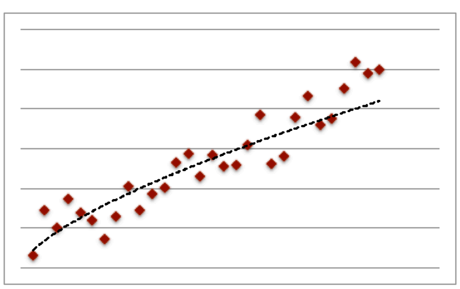


Figure 6.7a: Correct generalisation of noisy data

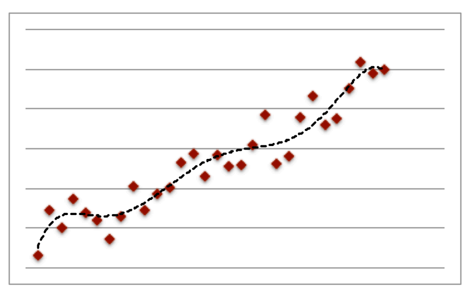


Figure 6.7b: Overfitting of a trend line

Overfitting can occur when there are too many parameters in a model that is being applied to observed data, or where there are too few observations. In this study the *railway-controlled vehicular private or staff* category of level crossings has a history of only two collisions in the SMIS database. For this category it is not possible to reliably infer general trends in the data and there is any regression model will necessarily overfit the data.

In general, level crossing collisions are – fortunately – rare events. Evans and Hughes (2019) state that in some cases fatalities occur only at a rate of one per five billion vehicle traverses. Therefore random events – even rare random events – can have an impact on the overall accuracy of any model. It is possible to imagine a road user who is in some way impaired and driving unsafely. Such a road user may have a high likelihood of being involved in an accident at any point during their journey; by chance the accident occurs at a level crossing. The accident would correctly be recorded as a level crossing collision, however the underlying cause of the collision may not be related to the presence of the level crossing nor the operation of a train at the level crossing, it was simply an accident waiting to happen that happened at a level crossing. Similarly a random failure of a road vehicle that leads to the vehicle becoming stranded on a level crossing could again lead to a level crossing collision that is nonetheless simply a random occurrence.

If collisions at level crossings were frequent events then occasional random events would not have a significant impact on general trends that would be inferred from regression analyses. However the relative infrequency of level crossing collisions makes all analyses sensitive to random events and can lead to overfitting when performing regression analysis. A possible method to improve the generality of models derived by regression analysis may be to inflate the data set by using additional data. As well as basing a regression model on observed collisions, it may be possible to include near-collision events in the analysis. A near-collision might be considered to have occurred when a road user was occupying a level crossing until only a few seconds before the arrival of a train.

Obstacle detectors are an emerging technology that can be used to detect the presence of obstructions on level crossings and are employed at some locations on the GB railway with the intention of preventing collisions at level crossings (Ohta, 2005; Hisamitsu *et al.*, 2008; Fakhfakh *et al.*, 2010). However it is not clear how this ambition can necessarily be achieved in accordance with railway operations. A high-speed train can take more than a minute to come to a stop, whereas a road user can move onto a level crossing and be clear again in only a few seconds. There could be a severe impact on the efficiency of the railway if all trains were required to take immediate action – perhaps commencing full emergency braking – if any obstacle were detected within their braking distance. The impact on the railway would be worse if members of the public were to vexatiously place obstacles on a level crossing as trains approach. It is clear that the ability of obstacle detectors to prevent immediate collisions is limited by operational requirements to be able to operate trains reliably at high speed.

Nevertheless an obstacle detector can identify when there has been an obstacle on a level crossing, even if the obstacle was cleared before the arrival of a train. It would be possible to determine the time between the obstacle clearing the level crossing and a train arriving. In this way obstacle detectors could be used to identify near-collisions in a way that could usefully inform the development of SRPTs. Inclusion of these data would create a more robust model than could be possible with collision data alone.

6.4 Inclusion of additional parameters and predictive accuracy testing

This study has considered the effect of road traffic volume on level crossing safety risk and sought to determine the degree to which the predictions of traffic models correlate with observed collisions. Naturally there may be factors other than only road traffic volume that affect level crossing safety risk. The limit on the scope of this study was imposed out of necessity since details of many of the SRPTs is not available, however details of each of the traffic models is available. Furthermore there is no general theory of level crossing safety risk. the various literature report on different factors that may affect risk but there is no general model that lists all the factors that affect safety risk.

A number of descriptive models have been created which correlate in varying degrees with observation; some of the models correlate very well (Table 6.1). In accordance with the earlier work by Evans and Hughes, the descriptive models consider the category of warning at each level crossing (passive, automatic, or railway-controlled) and the accessibility of the level crossing (whether public, or private or railway staff). However it cannot be expected that these are the only factors that affect safety risk; it is expected that there are other factors that also influence the likelihood of collisions. The derived traffic models described above could form the basis of a more general descriptive model. Such a model could be created by performing a multivariate regression analysis on various factors of level crossings. Alternatively, machine learning techniques are an emerging technology (for example Kelleher and Kelleher, 2019; Russell and Norvig, 2016) which can be used to detect patterns in data sets. For example machine learning clustering methods could be used to identify the factors that are most common amongst level crossing where more collisions or near-collisions occur compared with those where fewer collisions or near-collisions are experienced. Such an approach is consistent with the emerging approaches that are being applied more generally for railway safety management (Van Gulijk *et al.*, 2018; Hughes *et al.*, 2019; Van Gulijk *et al.*, 2016).

The predictive accuracy of an SRPT is the degree to which the safety risk predicted by the tool correlates with future collisions or near-collisions. A simple test of predictive accuracy could be conducted by using an SRPT to calculate at some point in the past (perhaps five years prior to the date of the assessment) what risk predictions would have been given for a range of level crossings. Such a study could then consider the collisions that have occurred during that time period and assess the degree to which the risk predictions and observed collisions correlate.

If it were possible to obtain evidence of the methods of calculation used by the various SRPTs, together of results of tests of predictive accuracy of each tool, it might be possible to determine which characteristics of a level crossings contribute to safety risk. It would therefore be possible to use the empirical results to inform an overarching academic theory of level crossing safety risk.

6.5 Discussion of results

The analysis described in this section has taken a fairly coarse approach to determine an approximate equation for the cumulative probability of a collision over road traffic volume. The approach taken is based on the approach used in the Kolmogorov Smirnov test to cumulate the values in the test variable as a means of addressing overdispersion in the data. The method of fitting a power curve ($y = ax^b$) to the distribution is not based on any scientific theory, rather it was based on a requirement to fit a curve that would approximately match the data: there is no *a priori* reason to presume that there would be a good fit between the data and a power curve. However the high R^2 values in Table 6.1 (up to 0.952) show that some of the curves correlate well with the observed rate of collisions. With the current understanding of the nature of safety risk at level crossings, this high degree of correlation can only be considered to be serendipitous. Whilst it is possible that the high degree of correlation is the consequence of an underlying phenomenon, at this stage no theory exists to explain why such correlation should exist. It is to be noted that the high degree of correlation indicated by the R^2 scores is not seen in all cases: Table 6.1 shows that the smallest degree of correlation is only 0.551.

Most notably, however, it can be seen that in every case (Figures 6.1 to 6.6) the cumulative number of collisions reduces as V increases. This result is shown mathematically in the results in Table 6.2 which are all presented in the form *CP(collisions per train traverse per day)* $= a'V^{b'}$. In every case the value of b' is below zero. Put simply, these results provide a compelling finding that the more road users who use a level crossing, the fewer collision occur per traverse.

This finding has not previously been demonstrated in the literature but has important implications for the railway industry in terms of the policy that is currently being followed by the GB railway infrastructure manager to close level crossings wherever possible (Network Rail, 2015). However when a level crossing is closed, it cannot be assumed that road users will abandon any desire to travel to the other side of the rail, rather road users will traverse the rail at other places. In some cases the other places may be grade separated crossings – bridges or tunnels – however in some cases the other place may be another level crossing. It is not clear that causing road users to divert from one level crossing to another will necessarily reduce the overall number of collisions; in fact it could potentially lead to more collisions. However the results of this study show that level crossing closure can be a valuable tactic as part of an overall programme of works to improve level crossing safety. For illustration, consider three vehicular level crossings each of the same category (say automatic public), two that have identical numbers of vehicular road users per day (say 5000) and the third that had exactly twice the number of road users per day as the other two (being 10,000). The third level crossing (with V equal to 10,000) will generally have a smaller number of collisions per traverse, and therefore will have a smaller number of total collisions than the other two combined.

6.6 Contribution

The following contributions to current knowledge have been made in this chapter:

Contribution 11: The study described in this chapter has identified that the distribution of collisions over road traffic volume (V) does vary in accordance with a power law of the form *collisions per train traverse per day* $= a'V^{b'}$. In every case the value of b' is below zero.

Contribution 12: The corollary of the observed power law indicates that level crossing closure can be considered an effective method to reduce the total number of collisions. This finding is particularly significant as the infrastructure manager of the GB railway has been undertaking a programme of level crossing closure although, to date, no studies have been undertaken to indicate that such an approach can be expected to improve safety overall.

Chapter 7: Machine learning of characteristics that affect level crossing safety

7.1 Machine learning

7.1.1 The need for machine learning methods

The studies in Chapters 3 to 6 have considered the degree to which traffic models correspond with observed collisions at level crossings in Britain. No evidence can be found of modern machine learning methods being applied in the determination of SRPTs, whereas it appears these methods could be useful as they are perfectly suited to discovering complex, non-linear correlations between *features* (such as the characteristics of level crossings) and *labels* (such as observed collisions).

The data published by Network Rail provides road user census data for all level crossings in Britain and it is therefore possible to analyse the degree of correspondence between the various traffic models and observed collisions. Since the method of calculation of the SRPT used in Britain (the ALCRM) has not been published, it has not been possible to assess the degree of correspondence between the results of the SRPT and observed collisions.

In addition to the data on road user volumes at each level crossing, data Network Rail also gives additional information listed in Table 7.1. In particular data are provided on operational characteristics, hazards and risk controls at each level crossing. It is possible to perform a test to determine if there are significant correlations between the numbers of collisions at a level crossing and these additional features.

Table 7.1: Data on level crossings provided by Network Rail (2017)

Category	**Data provided by Network Rail**
Identification information	Unique number; level crossing name; class of level crossing
Location	Latitude and longitude, location description, railway engineers line reference; postcode
Operational characteristics	Types of train (freight and passenger); rail speed; number of trains per day; numbers of road users (vehicular and pedestrian)
Hazards	Items from the following list: *blocking back; crossing approach; crossing is near a station; deliberate misuse or user error; frequent trains; gates open; infrequent trains; large numbers of HGVs; large numbers of users; low sighting time;* no specific risk drivers identified*; poor visibility for approaching road vehicles; sun glare*.
Risk controls	Items from the following list: *audible alarm; barrier; CCTV monitoring by signaller; full barrier equipment; gates; half barrier equipment; road markings; road traffic light signals; signage; stop boards provided on the train approaches - trains stop and drivers sound the train horn before proceeding; telephones provided for vehicle users; train signalling protection; whistle boards provided on the rail approach in one direction - train horn audible warning given (06:00 to 23:59); whistle boards provided on the rail approaches - train horn audible warning given (06:00 to 23:59)*.
Safety-related events	Numbers of *accidents*, *incidents*, and *near misses*
Risk assessment details	Date of the most recent assessment and due date of the next assessment, risk scores in two categories: individual risk ranked from A to M, and collective risk ranked 1 to 13

Table 7.2 shows example shows the data for four level crossings selected from the dataset.

Table 7.2: Examples level crossings showing values; data taken from Network Rail (2017)

Identification number	2144	6973	4528	3297
Name	Crewkerne	Huish	Stowgate	Church Dam
Level crossing class	Public Highway Automatic Half Barriers	Public Highway Manned Barriers CCTV Monitored	Public Highway Automatic Half Barriers	Public Highway User worked Crossing
Latitude co-ordinates	50.872099	51.376721	52.678152	52.563754
Longitude co-ordinates	-2.79278	-2.862732	-0.24541	1.577176
Location	South Somerset District	Puxton CP	Deeping St. James CP	Reedham CP
Engineers line reference	BAE2-132.0073	MLN1-132.0242	WEB0-084.0835	RBY0-012.1669
Postcode	TA188PF	BS246RZ	PE6 8RW	NR133UE
Types of trains	Passenger & Freight	Passenger & Freight	Passenger & Freight	Passenger & Freight
Line speed (mph)	75	100	75	60
No. of trains per day	37	100	28	4
Vehicular road users per day	1296	189	68	1
Pedestrians or cyclists per day	81	Infrequent	14	Infrequent
Key risk drivers	Poor Visibility for Approaching Road Vehicles	Large Numbers of HGVs	No Specific Risk Drivers Identified	Infrequent trains; Sun Glare
Risk controls	Half barrier equipment; Road traffic light signals; Audible alarm; Signage	Train signalling protection; CCTV monitoring by signaller; Full barrier equipment; Road traffic light signals; Audible alarm; Signage	Half barrier equipment; Road traffic light signals; Audible alarm; Signage	Telephones provided for vehicle users; Gates; Signage
Near Miss history	Nil	Nil	Nil	Nil
Incident history	1	1	2	2
Accident history	Nil	Nil	Nil	Nil
Current assessment date	Mar-16	Sep-15	May-14	Jun-15
Next assessment due date	Jun-18	Dec-18	Aug-17	Aug-18
Individual risk letter	F	I	E	D
Collective risk number	4	8	6	8

The data describe all level crossings in Great Britain from those in rural areas with low traffic volumes, to those on heavily used public roads in urban areas. Table 7.3 shows the range of values for the numerical data for operational characteristics in the dataset.

Table 7.3: Range of values for numerical data on operational characteristics in Network Rail (2017)

Characteristic	Lowest value	Highest value
Trains per day	1	479
Vehicles per day	*Infrequent*	29,592
Line speed (mph)	5	125

As noted in Section 4.1.1, where the data show that there are *infrequent* road users, the values of 0.5 road users per day was used in this analysis. The level crossings also have a range of different hazards and risk controls. Some level crossings have no information on hazards, in these cases the entry *no specific risk drivers* is recorded in the dataset. The largest number of hazards recorded for a level crossing was seven, being:

- poor visibility for approaching road vehicles;
- crossing is near a station;
- crossing approach;
- large numbers of users;
- sun glare;
- deliberate misuse or user error; and
- blocking back.

In all cases, level crossings have some signage, therefore all level crossings in the dataset have at least one entry for the control devices. The largest number of warning devices for a level crossing was six, being:

- signage;
- train signalling protection;
- CCTV monitoring by signaller;
- full barrier equipment;
- road traffic light signals; and
- audible alarm.

In principle it would be possible to perform a statistical regression analysis to determine the degree of correlation between any of these characteristics and the rate of collisions at level crossings. However such an analysis would be unwieldy because of the large number of characteristics any of which by itself, or in combination with other characteristics may impact on safety risk. For example considering the hazards identified in Table 7.1, it is possible that by itself large numbers of HGVs (heavy goods vehicles) does not have a significant impact on safety risk, however in conjunction with low sighting time, the impact of the combination of hazards may be significant. Similarly there may be risk controls which are effective only when applied in combination with other controls. Furthermore is possible that some risk controls – or combinations of risk controls – are effective at controlling on some hazards, or combinations of hazards.

The data from Network Rail identifies 12 different hazards (ignoring the option *no specific risk drivers*). This list allows for a large number of combinations of hazards to be described at any level crossing. The total number of combinations is given by summing the binomial coefficients for the number of ways (C) of selecting k items from a list of n options for all values of k from zero to n. Equation 7.1 shows the method of calculating the binomial expansion.

$$C = \sum_{k=0}^{n} \frac{n!}{k!\,(n-k)!} \qquad \text{Equation 7.1}$$

Where:

C	is the total number of combinations;
k	is the number of items to be selected from the list of hazards;
n	is the number of items in the list of hazards.

Since the data from Network Rail identifies 12 hazards, putting the value n=12 into Equation 7.1 gives a result of $C_{hazards}$ = 4096. Similarly the data identify 14 different risk controls which, again, could occur in any combination at a level crossing. Putting the value of n=14 into Equation 1 gives a result of $C_{controls}$ = 16,384. Conceivably, any combination of risk controls may be effective in reducing safety risk when used with any combination of hazards. As such the total number of cases to be considered from these data are:

$$C_{total} = C_{hazards} \times C_{hazards} = 4096 \times 16{,}384 = 67{,}108{,}864$$

Since there are data only 3742 level crossings, clearly it is not necessary to test every possible combination of hazards and risk controls. Nevertheless there remains a large number of combinations that would be need to be tested during a regression analysis to determine with confidence whether there are specific combinations of hazards and risk controls that have a significant impact on safety risk.

A further complication for performing a regression analysis is that the relationship between the characteristics and safety risk may not be linear. The studies described in Chapter 6 demonstrates that where there is a relationship between road traffic volume and safety risk the covariance does not follow a linear pattern, rather the relationship is better described by the formula: $CP = a'V^{b'}$. Since no theory has been presented for the possible relationships between characteristics and collision rates, the nature of this study is *exploratory* to determine if any plausible relationship can be identified. As such there are countless possible types of relationship that could be tested for correlation with observation.

Conversely, a number of the machine learning methods are not expected to be exhaustive in their pattern searches, a number of the methods, such as *random forests*, are purposefully designed to generate approximate patterns that avoid overfitting the data. Such a heuristic approach substantially reduces the computational effort required in analysis and therefore lends itself to the type of exploratory analysis being undertaken in this study. The method applied in this study

7.1.2 Full and fractional factorial design

When designing experiments where there are a large number of input factors, *full factorial design* is a technique that is often used in chemical processes (Patience, 2017) and medical research (Beg and Hasnain, 2019). Full factorial design reduces the number of experiments that are needed by measuring the effects of input factors at only the extremes of the range their values: both at the extreme low end and the extreme high end of the range. Using this approach, it is assumed that the effect of each factor on the output at intermediate values can be interpolated from the values at the extreme ends. Using full factorial design, each factor and each combination of factors are tested. Where there are a large number of factors, the numbers of combinations can still be large and this approach can still require a large number of experiments.

Fractional factorial design further reduces the number of experiments required by assuming that the effects of some combinations of factors can be modelled by considering the effects of other combinations: for example the effects of factors A, B, and C on the output, can be inferred by examining the effects of factors A and B and also the effects of factors A and C on the output. Fractional factorial design allows the number of experiments to be reduced by around 50% (Sarker and Nahar, 2018).

It may appear that full or fractional factorial design could be a useful approach for constructing machine learning experiments to reduce the number combinations of factors that need to be considered. However there are some important considerations: firstly is that the approaches of full and fractional factorial design assume that the effects of factors can be modelled by considering their influences at the extreme ends of their ranges. In effect there is an assumption that the effect of factors on the output is approximately linear: if the effect of a factor at the higher end of its range produces a positive effect on the output, then it can be assumed in general that larger concentrations of the factor will always produce a positive effect. By considering the effect at only the ends of the range this approach to experimental design cannot consider non-linear effects that occur at intermediate values. However, as discussed in Sections 2.15 and 7.1.1, it is possible that there may be non-linear effects with specific factors, or combinations of factors, that affect safety risk at level crossings. Furthermore fractional factorial design attempts to reduce the number of experiments by assuming that some combinations of factors do not need to be modelled as the effects of specific combinations can be inferred by considering other combinations. Such an approach is valid where the effects of particular combinations of a factor are well understood; perhaps as a result of prior experiments which have demonstrated that particular factors are independent of each other. However in the study of level crossing safety there is a dearth of evidence of the effects of any combinations of factors on safety risk. As discussed in Section 2.2, a number of researchers have considered the effects of individual factors on level crossing safety risk, but there is not sufficient evidence on the effects of combinations of factors to be able to confidently assume that some factors are mutually independent in their effect on level crossing safety risk.

Secondly, fractional factorial design is typically applied in cases where there are at most dozens of experiments to be performed (Sarker and Nahar, 2018). For the case of machine learning considered in this study there are more then 67 million combinations of factors to be considered: even a 50% reduction in the number of experiments would still leave more than 33 million combinations to be considered.

Lastly, the approaches of full and fractional factorial design are meaningful for experiments that require physical recourses, such as experiments that consider the effects of different concentrations of chemical components on a specimen of biological material. In such experiments it is common that a factor, such as a chemical reactor, cannot be reused once it has been used in one experiment. In these cases the physical resources required may be expensive to obtain, or scarcity of materials may mean that there are insufficient quantities of materials to perform the full number of experiments. For this study, the inputs to the experiments are data stored in electronics computers. Modern computers are able to store large quantities of data very cheaply: the cost of copying and storing data to run additional experiments is virtually zero. The main limitation on the number of experiments that can be performed is the time required for computation. Modern computers are able to perform many millions of computations in a short amount of time, and the various machine learning methods use a number of heuristics to reduce the computational space that needs to be searched in performing calculations. The methods used to reduce the number of experiments required for performing physical experiments on physical resources are not necessarily well-suited for addressing the requirements of data-driven experiments.

7.2 Method of application

Machine learning methods were applied to determine whether a correlation could be identified between the features of a level crossing and the history of collisions. Data on level crossing characteristics from the Network Rail spreadsheet were used as in the preceding work described in Chapters 4 to 6; collision data from the SMIS database was used. The level crossing characteristic data used in the analysis were all data fields from the following categories shown in Table 7.1:

- category of level crossing;
- operational characteristics;
- hazards; and
- risk controls.

Where numeric values were provided for the features, these values were used directly in the machine learning process. A *one-hot encoding* was used to describe the hazards and risk controls, *i.e.* a separate column was added to the data for each type of hazard and for each type of risk control. A value of one was entered in the column if the hazard or risk control was present at a level crossing, otherwise the value zero was stored. Therefore each level crossing had at least one non-zero entry for risk controls, since all level crossings have at least signage even if they have no other risk controls. A column was not added for the entry *no specific risk drivers* since this can be inferred from the absence of any non-zero values in any of the columns describing hazards.

In keeping with the nomenclature described in Chapter 2, these data were used as the *features* for each level crossing, collision data from the SMIS database were used as the *labels*. In Chapters 4 to 6, the collision data were normalised by *collisions per train traverse per day*. Ideally it would be desirable for a machine learning model to be able to correctly predict collision rates for a level crossing. For the initial exploratory study it is not clear which, if any, of the machine learning methods would be well-suited to identify patterns of correlation. Instead level crossings were labelled with a Boolean value indicating whether each level crossing had a history of any collisions. The test was therefore to determine if the characteristics of a level crossing from the Network Rail spreadsheet were adequate to accurately determine level crossings where collisions have occurred. In the input data, a value of zero was used to indicated no history of collisions, and one was used to indicate that at least one collision had been recorded.

These data were saved in a tab-separated values file. The file consisted of 38 columns. The structure of the data is shown in Table 7.4.

Table 7.4: Structure of data used in the machine learning analysis

Columns	Data	Details
1	Unique identification number for each level crossing.	–
2	Description of level crossing class.	See entry for Column 3 below.
3	Description of level crossing class, and a numeric identifier for each class.	The classes and numeric identifiers are: 5: Passive Vehicular Public 6: Passive Vehicular Private or Staff 7: Automatic Vehicular Public 8: Automatic Vehicular Private or Staff 9: Railway-controlled Vehicular Public 10: Railway-controlled Vehicular Private or Staff
4	Operational characteristic: number of trains per day	Numeric value between 1 and 479.
5	Operational characteristic: maximum train speed	Numerical value between 5 and 125 (mph).
6	Operational characteristic: speed difference between the up and down railway lines.	Numerical value between 0 and 50 (mph)
7 and 8	Operational characteristics: whether the railway line is used by passenger and / or freight trains.	One-hot encoding with values of zero or one for passenger trains (Column 7) and freight trains (Column 8).
9	Operational characteristic: number of vehicles traversing the level crossing per day.	Numerical value between 0.5 (infrequent) and 29,592
10	Operational characteristic: number of pedestrians traversing the level crossing per day.	Numerical value between 0 and 30,051.
11 to 23	Hazards	One-hot encodings for each of the hazards: • blocking back • crossing approach • crossing is near a station • deliberate misuse or user error • frequent trains • gates open • infrequent trains • large numbers of HGVs • large numbers of users • low sighting time • no specific risk drivers identified • poor visibility for approaching road vehicles • sun glare

Columns	Data	Details
24 to 37	Risk controls	One-hot encodings for each of the risk controls: • audible alarm • barrier • CCTV monitoring by signaller • full barrier equipment • gates • half barrier equipment • road markings • road traffic light signals • signage • stop boards provided on the train approaches - trains stop and drivers sound the train horn before proceeding • telephones provided for vehicle users • train signalling protection • whistle boards provided on the rail approach in one direction - train horn audible warning given (06:00 to 23:59) • whistle boards provided on the rail approaches - train horn audible warning given (06:00 to 23:59)
38	Label indicating whether the level crossing had a history of collisions.	Boolean value

The file had 3742 rows: one for each vehicular level crossing.

The tests were performed using a software program written in the Python programming language; this language was selected because of the extensive range of software libraries available and in particular the widely used *sci-kit learn* machine learning library (Pedregosa *et al.*, 2011). The code used in the tests is shown in Appendix B.

The code contains variables that select which features will be used in the analysis and which machine learning method will be applied. The code allowed the user to select groups of features to be included in the analysis, for example a group of all features describing hazards, or all features describing risk controls. Alternatively features could be selected singly or in any combination, or all features could be included in the analysis.

Following from the review of machine learning methods described in Section 2.15, each test was performed with one of the selected machine learning methods for the choice of: decision tree method; random forest; ANN; or support vector machine as shown in Table 7.5.

Table 7.5: Machine learning methods and *sci-kit learn* configuration used in the study

Method	sci-kit learn library module	Model configuration
Decision tree	`sklearn.tree.DecisionTreeClassifier`	*n/a*
Random forest	`sklearn.ensemble.RandomForestClassifier`	`max_depth: 5` `n_estimators: 10`
Artificial neural network (multi-layer perceptron)	`sklearn.neural_network.MLPClassifier`	`alpha: 1` `max_iter: 1000`
Support vector machine (Gaussian radial basis function)	`sklearn.svm.SVC`	`gamma: 2` `C: 1`

7.2.1 Recall value

A significant proportion of level crossings in Britain have no history of collisions. Table 7.6 shows the proportion of vehicular level crossings where no collisions have been recorded.

Table 7.6: Count of level crossings in each class with no history of collisions

Class of level crossing	Total number of level crossings	Number of level crossings with no history of collisions	Proportion of level crossings with no history of collisions
Passive Vehicular Public	112	102	91.1%
Passive Vehicular Private or Staff	2114	2040	96.5%
Automatic Vehicular Public	579	475	82.0%
Automatic Vehicular Private or Staff	112	88	78.6%
Railway-controlled Vehicular Public	774	760	98.2%
Railway-controlled Vehicular Private or Staff	51	49	96.1%
Total	3742	3514	93.9%

A weighted average of these results shows that 93.9% of vehicular level crossings have no history of collisions. As such it would be possible to create a model that predicted for any level crossing that there would be no collisions and the model would have an overall accuracy of 93.9%. Such a result would not provide a valuable tool for risk management on the railway. To address this issue, *recall* was used as the test of each model's performance, the recall value is given as: $recall = \frac{count\ of\ true\ positives}{true\ positives + false\ negatives}$

7.2.2 Selection of training set size

When performing supervised machine learning there is a question regarding the proportion of records that should be used to train the model (the *training set*) and the proportion that are used to validate the model (the *testing set*). If the training set is too small, the machine learning method cannot sufficiently learn the characteristics of the data and will provide inaccurate results. Conversely, too large a training set will lead to over-fitting and will reduce the generality of the resultant model.

The literature commonly show cases where the testing sets is 20%, 30% or 40% of the data set, *for example* Stamp (2017), Alpaydin (2009) and Raschka (2015). In general, however, little justification is given for the selection of any particular proportion.

Amari *et al.* (1997) proposed that the optimal proportion of testing records is given by:

$$r'_{opt} = \frac{1}{\sqrt{f}} \qquad \text{Equation 7.2}$$

Where:

r' denotes the testing set, and r'_{opt} is the optimal value to maximise the learning and minimise over-fitting;

f is the number of features in the data.

A proof of this formula is also given in Guyon (1997). For the data used in this study, there are 33 features, being: 7 operational characteristics; 12 values describing hazards; and 14 values describing risk controls. Applying Equation 7.2 to this data set gives:

$$r'_{opt} = \frac{1}{\sqrt{33}} = 0.174 = 17.4\% \qquad \text{Equation 7.3}$$

Another point to note is that the data used in this study are *sparse*: most of the data describe negative examples (level crossings where there is no history of collisions). To provide a large enough sample of positive examples to allow the models to produce accurate results, a large training set will be required; therefore a larger training set of the order 80% is again supported.

Since the training set is a random selection of records from the full data set, it is possible that divisions of the data will lead to cases where there are too few positive examples in the training set, or too few positive examples in the testing set. To overcome this problem that may occur with a single application of the method, each method was repeated 1000 times and the results of each test were combined to produce an overall recall value. Another approach that can be used to avoid overfitting is to reduce the number of features in the data (Guyon, 1997; Alpaydin, 2009). This study uses an exploratory approach that identifies the methods that produce the best recall values and then applies a reductionist approach to identify which features or sets of features produce the best recall values.

Finally it is noted, that the effect on the accuracy of the results does not necessarily vary dramatically when varying the size of the training set. Towards Data Science (2020) identified that a 7% variation in the size of the training set produce only a 2% increase in the precision, recall and f1-scores. Based on these findings, a training set of 80% was used for this study.

7.2.3 Iterative application

For each of the methods, tests were run for vehicular level crossings with the following sets of features:

- operational characteristics
- hazards
- risk controls
- all of the above, *viz.* operational characteristics, hazards, and risk controls

For four methods with four sets of characterises gives a total of 16 tests. The results are shown in Table 7.7. These initial results showed that the decision tree was the only method to produce recall values of more than 20%: the recall method of the other methods was in every case less than 4% and in many cases zero. Since the decision tree method appeared to be the most effective, further tests were performed using this method with smaller sets of features to determine whether particular features correlate with good recall performance. Tests were run for each class of level crossing for each category of features (operational characteristics, hazards, or risk controls). As with the initial results, there was a clear distinction in the results: tests performed with operational characteristics as features of the data produced recall results of up to 31%, whereas only one of the tests that used hazards or risk controls as features produced a non-zero result of only 1.62%. The tests were again repeated using the decision tree method, with each of the operational characteristics tested individually. The results of all these tests are shown in Table 7.8.

7.3 Recall values from machine learning tests

The recall values for each of the tests are shown in the following tables. Each row of the tables shows the mean recall results averaged over 1000 executions of the test. Table 7.7 shows the results of the initial tests; the five columns of the table show:

- *test number*: a unique identifier for each test;
- *machine learning method*: which one of the four methods was used in the test;
- *classes of level crossing*: the results shown in Table 7.7 are for the initial analysis which was applied to all classes of level crossing;
- features: which of the features (operational characteristics, hazards, controls, or all features) where used in the analysis to test for correlations; and
- *average recall value*: the mean value of the recall value from 1000 tests.

Table 7.7: Results of initial tests

Test	Machine learning method	Classes of level crossing	Features	Average recall value
1	Decision tree	All	operational characteristics	19.97%
2	Decision tree	All	hazards	0.00%
3	Decision tree	All	controls	0.00%
4	Decision tree	All	all	22.03%
5	Random forest	All	operational characteristics	0.00%
6	Random forest	All	hazards	0.00%
7	Random forest	All	controls	0.00%
8	Random forest	All	all	0.08%
9	Artificial neural network	All	operational characteristics	3.06%
10	Artificial neural network	All	hazards	0.00%
11	Artificial neural network	All	controls	0.00%
12	Artificial neural network	All	all	3.78%
13	Support vector machine	All	operational characteristics	0.14%
14	Support vector machine	All	hazards	0.00%
15	Support vector machine	All	controls	0.00%
16	Support vector machine	All	all	0.15%

The results in Table 7.7 show that only the decision tree method produces an average recall value above 4%, so this method was used for further study. Table 7.8 shows the results of the tests using only the decision tree method. The column in the table show:

- *test number*: a unique number for each test, continuing the numbering from Table 7.7;
- *classes of level crossing*: which class of level crossings the test was applied to;
- *features*: the groups of features, or individual features that were used in the test; and
- *mean recall value*: the mean value of the recall value from 1000 tests.

Table 7.8 Results of detailed tests

Test number	Classes of level crossing	Features	Mean recall value
17	Passive Vehicular Public	operational characteristics	31.61%
18	Passive Vehicular Private or Staff	operational characteristics	11.87%
19	Automatic Vehicular Public	operational characteristics	25.22%
20	Automatic Vehicular Private or Staff	operational characteristics	27.94%
21	Railway-controlled Vehicular Public	operational characteristics	2.77%
22	Railway-controlled Vehicular Private or Staff	operational characteristics	0.30%
23	Passive Vehicular Public	hazards	0.00%
24	Passive Vehicular Private or Staff	hazards	0.00%
25	Automatic Vehicular Public	hazards	0.00%
26	Automatic Vehicular Private or Staff	hazards	0.00%
27	Railway-controlled Vehicular Public	hazards	0.00%
28	Railway-controlled Vehicular Private or Staff	hazards	0.00%
29	Passive Vehicular Public	controls	0.00%
30	Passive Vehicular Private or Staff	controls	0.00%
31	Automatic Vehicular Public	controls	0.00%
32	Automatic Vehicular Private or Staff	controls	1.62%
33	Railway-controlled Vehicular Public	controls	0.00%
34	Railway-controlled Vehicular Private or Staff	controls	0.00%
35	Passive Vehicular Public	Trains per day	47.12%
36	Passive Vehicular Public	Max line speed	0.13%
37	Passive Vehicular Public	Speed difference up/dn	0.00%
38	Passive Vehicular Public	Train type: Passenger	0.00%
39	Passive Vehicular Public	Train type: Freight	0.00%
40	Passive Vehicular Public	Vehicles per day	6.79%
41	Passive Vehicular Public	Pedestrians per day	6.06%
42	Passive Vehicular Private or Staff	Trains per day	0.16%
43	Passive Vehicular Private or Staff	Max line speed	0.00%
44	Passive Vehicular Private or Staff	Speed difference up/dn	0.00%
45	Passive Vehicular Private or Staff	Train type: Passenger	0.00%
46	Passive Vehicular Private or Staff	Train type: Freight	0.00%
47	Passive Vehicular Private or Staff	Vehicles per day	4.72%
48	Passive Vehicular Private or Staff	Pedestrians per day	2.19%

Test	Classes of level crossing	Features	Mean recall value
49	Automatic Vehicular Public	Trains per day	9.57%
50	Automatic Vehicular Public	Max line speed	6.52%
51	Automatic Vehicular Public	Speed difference up/dn	0.00%
52	Automatic Vehicular Public	Train type: Passenger	0.00%
53	Automatic Vehicular Public	Train type: Freight	0.00%
54	Automatic Vehicular Public	Vehicles per day	17.80%
55	Automatic Vehicular Public	Pedestrians per day	4.83%
56	Automatic Vehicular Private or Staff	Trains per day	19.43%
57	Automatic Vehicular Private or Staff	Max line speed	9.04%
58	Automatic Vehicular Private or Staff	Speed difference up/dn	0.00%
59	Automatic Vehicular Private or Staff	Train type: Passenger	0.00%
60	Automatic Vehicular Private or Staff	Train type: Freight	0.00%
61	Automatic Vehicular Private or Staff	Vehicles per day	27.76%
62	Automatic Vehicular Private or Staff	Pedestrians per day	8.18%
63	Railway-controlled Vehicular Public	Trains per day	0.00%
64	Railway-controlled Vehicular Public	Max line speed	0.00%
65	Railway-controlled Vehicular Public	Speed difference up/dn	0.00%
66	Railway-controlled Vehicular Public	Train type: Passenger	0.00%
67	Railway-controlled Vehicular Public	Train type: Freight	0.00%
68	Railway-controlled Vehicular Public	Vehicles per day	0.00%
69	Railway-controlled Vehicular Public	Pedestrians per day	0.00%
70	Railway-controlled Vehicular Private or Staff	Trains per day	0.00%
71	Railway-controlled Vehicular Private or Staff	Max line speed	0.00%
72	Railway-controlled Vehicular Private or Staff	Speed difference up/dn	0.00%
73	Railway-controlled Vehicular Private or Staff	Train type: Passenger	0.00%
74	Railway-controlled Vehicular Private or Staff	Train type: Freight	0.00%
75	Railway-controlled Vehicular Private or Staff	Vehicles per day	0.00%
76	Railway-controlled Vehicular Private or Staff	Pedestrians per day	0.00%

The recall values for Tests 35 to 76 shown in Table 7.8 are repeated in Table 7.9 which sets out the results in a different tabular format to allow easier comparison of results by features and classes of level crossing.

Table 7.9: Summary of recall values from rows 35 to 76 of Table 7.8

	Class of level crossing					
Individual characteristic	**Passive vehicular public**	**Passive vehicular private or staff**	**Automatic vehicular public**	**Automatic vehicular private or staff**	**Railway-controlled vehicular public**	**Railway-controlled vehicular private or staff**
Trains per day	47.12%	0.16%	9.57%	19.43%	0.00%	0.00%
Max line speed	0.13%	0.00%	6.52%	9.04%	0.00%	0.00%
Speed difference up/dn	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Train type: Passenger	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Train type: Freight	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Vehicles per day	6.79%	4.72%	17.80%	27.76%	0.00%	0.00%
Pedestrians per day	6.06%	2.19%	4.83%	8.18%	0.00%	0.00%

7.4 Interpretation of results

Considering the results of the initial tests in Table 7.7., nine of the 16 rows show a recall value of 0.00%. The mean value of all the results in the table is only 7.03%. Only two tests obtained a recall value above this mean: both of these results were obtained when the decision tree method was applied. These two results were obtained when the *features* of the tests were either operational characterises, or all characteristics. The tests described in Table 7.7 were purposefully broad in their scope as an initial exploration of the methods and therefore provide little insight into the correlation between specific characteristics of level crossings and collision rates. Rather the tests indicate that if any correlation can be found using machine learning methods then it is likely that the decision tree method is best suited to the data set to identify the correspondence. Table 7.8 shows the results of the further exploration of the data when this method was applied to more refined subsets of the data.

The results for tests 17 to 34 show the results of applying the decision method to each class of level crossing individually for each set of features, *viz. operational characteristics*, *hazards*, and *controls*. These results show that it is only the operational characteristics that consistently produce recall values more than 1.00%. This result indicates that there is a possibility that, for a specific class of level crossing, it is only the operational characteristics – and not the hazards nor control devices – that have any meaningful impact on level crossing safety. Such a finding would be significant as it is contrary to the assumptions underlying the creation of many SRPTs, some of which consider a large number of characteristics of level crossings including individual hazards and controls (Baker and Heavisides; 2007).

Prima facie the finding that there seems to be a correlation between operational characteristics and observed collision may appear to contradict the findings of the study in Chapter 5 which found that, in most cases, there was no statistically significant correlation between the traffic models and observed collisions. However it should be noted that the study described in Chapter 5 found that there was some degree of correlation, but that in all but one case, the degree of correlation was not sufficient to pass the Kolmogorov Smirnov test. The machine learning method applied in this study has shown, again, that there appears to be some correlation but, once again, the degree of correlation does not support any theoretical model. As such, these results serve as a cross-validation of the two methods: use of statistical methods and the Kolmogorov Smirnov test, and the machine learning methods. The most salient finding from these results appears to be that although the correlation between operational characteristics and observed collisions is weak, it is nevertheless the strongest result in the study. It is important to emphasise that this finding does not prove that there is no correlation between collision rates and the hazards and controls at level crossings, rather the finding shows that no correlation can be found, even when using the machine learning method that appears from initial tests (Table 7.7) to produce the largest recall values. For any complex system it is never possible to prove that no correlation exists between various sets of variables. It remains possible that there is some correlation within the data, albeit only a weak correlation, or perhaps a correlation with a theoretical model that has not yet been identified. However the machine learning methods used in this study do not require an *a priori* model for testing, rather the methods attempt to identify with any correlation can be found. As such, the lack of a significant results when a range of different machine learning methods has been applied provides some indication that there may be no meaningful correlation. This result is consistent with the observation that there is no evidence of any of SRPT producing results that correlate with observed collision rates (Chapter 2).

Another observation from the results for Tests 17 to 22 is that the recall values for railway-controlled level crossings are low compared with other classes of level crossing: below 3%, whereas the recall rates of other classes of level crossing are above 11% and update to 31%. It is often the case in statistical analysis or machine learning that low scores are obtained when there is only a small number of samples in the test data and therefore it is not possible to identify general trends that are true for all instances. However the small values for railway-controlled level crossings cannot be readily explain in this way for the data used in these tests. Figure 7.1 provides graphical representation of the numbers of level crossings in each class and the recall values obtained during Tests 17 to 22. Again, these results are not, by themselves, conclusive, however they lead to a hypothesis that it is possible that whilst operational characteristics have some correlation with collision rates, the effect is very much reduced for railway-controlled level crossings. It is noted that, in Britain, railway-controlled level crossings have full gates or booms covering the full width of the road: both the approach and departure carriageways. This type of barrier may, by itself, be the factor that leads to reduced collision rates at this class of level crossing (Evans and Hughes; 2019). As such it is possible that when full-width barriers are provided, other operational characteristics of level crossings do not have any meaningful bearing on the occurrence of collisions.

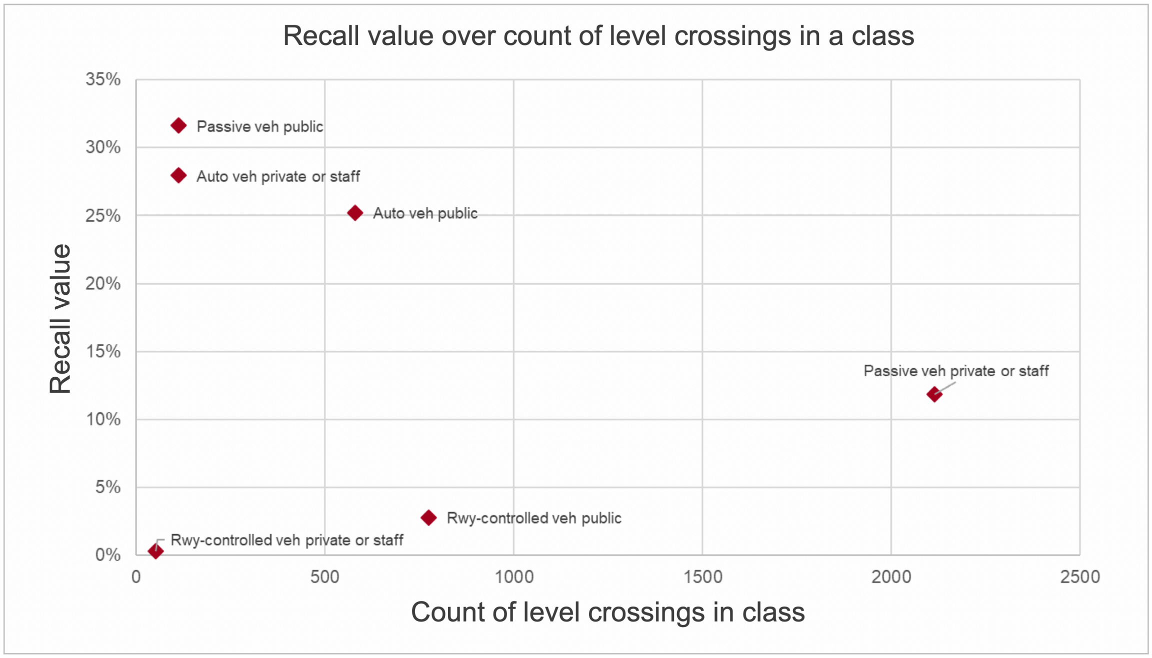


Figure 7.1: Recall value over count of level crossings in a class

Tests 35 to 76 consider the effects of individual characteristics for each class of level crossing, these data are repeated in Table 7.9 for ease of comparison between characteristics and classes of level crossing. These results show that the only individual characteristics that produce non-zero recall values are: the numbers of trains per day, the maximum line speed, and the numbers of road users. Within these results there is, again, a variety in the values, the non-zero values range from 0.13% to 47.12%, with little consistency in either the characteristics nor the classes of level crossings that are producing the higher or lower values.

The results in Table 7.9 show that, for some classes of level crossing, it is possible to use the decision tree method to identify some degree of correspondence between collisions at level crossings and some operational characteristics, specifically: the maximum line speed and the numbers of trains and road users per day. However it must be noted that the largest recall value is only 47.12% which. From the definition of the recall value, this result means that for level crossings where there is a history of collisions then – at best – only 47.12% of these level crossings can be correctly identified by considering the characteristics of a level crossing. In summary, the method is no better than a random coin-flip at correctly identifying level crossings where there has been a history of collisions. Furthermore this value is obtained only after testing a number of machine learning methods and selecting the one that consistently produces the largest recall values.

The conclusions that can be drawn from this result are that either:

1. there exists a correlation between the physical and operational characteristics of level crossings and rates of collisions, but the methods used were not sufficiently powerful to identify the correlation; or
2. there is a correlation but the characteristics recorded in the database are not the correct data to allow the effect to be identified; or
3. no meaningful correlation exists.

It is an underlying assumption in all SRPTs that there is a meaningful correlation between the characteristics of level crossings and collision rates, it is therefore necessary to consider the implications of these three possibilities.

The first possible possibility – that there is a correlation, but the methods used cannot properly identify it – raises the question of the value of the methods used to create the SRPTs currently in use around the world. Furthermore, the question should be asked whether the SRPTs currently in use are in any way effective at identifying level crossings where there is a high risk of collision. There various SRPTs around the world use a variety of methods to estimate safety risk and there is no evidence available that any of them produces results that correlate with observed collision rates. A further consideration is that the method that was used in this study uses powerful machine learning techniques that have been developed specifically for identifying correlations in data. The method used in this study has a number of advantages over the statistical techniques that have traditionally been used for identifying correlations in data. In fact the method used in this study is relatively modern (the sci-kit learn machine learning library was first published in 2007 and has been incrementally revised since) and post-dates the creation of a number of the SRPTs. Whilst it is never possible to conclude that no correlation exists, the rigorous and methodological approach used in this study, together with use of modern machine learning methods, provides a strong argument for the case that there may, in fact, be no correlation.

The second possibility is that the characteristics of a level crossing do affect the safety risk – and hence the rate of collisions – but that the data recorded in the Network Rail dataset do not describe these characteristics: perhaps there are other characteristics that affect the risk. This possibility therefore raises the question of what these other characteristics might be. The available data on level crossings describe the physical and operational characteristics of the level crossings including the hazards and control devices in place. These data have been collected by professional staff from the railway who have detailed operational knowledge of level crossings. The staff collecting the data are, arguably, the best qualified to identify the data that are relevant. If, in future, it is possible to obtain a broader range of data regarding level crossings, then it is possible that the study could be repeated. It should also be considered that physical and operational characteristics of level crossings may have only a small influence on level crossing safety risk: there may be other characteristics that have a larger affect. For example characteristics of road users – such as their skill in driving a vehicle, or whether they are distracted, intoxicated or otherwise incapacitated – may have a more significant influence on level crossing safety than its physical and operational characteristics.

The final possibility to be considered is that the class of the level crossing is the only factor that has a influence on safety risk (Evans and Hughes; 2019) and that there is no other meaningful correlation with other characteristics. This conclusion flies in the face assumptions that underlie the development of the many SRPTs in use around the world and may initially seem implausible. However the lack of correlation between characteristics may be a secondary effect that occurs as a result of road user behaviours. The theory of *risk homeostasis* (Wilde; 2014) states that humans adapt their behaviour to the perceived level of risk they face. When applied to level crossing safety, this theory proposes that if road users perceive a level crossing to have a high degree of safety risk, perhaps because of the presence of hazards or absence of risk controls, they will adapt their behaviour to reduce the risk. Currently, no published research can be found on risk homeostasis of road users at level crossings and the available data do not lend themselves to such a study. It would therefore be a matter for further study to investigate this question.

7.5 Contribution

The following contributions to current knowledge have been made in this chapter:

Contribution 13: The study described in this chapter has demonstrated a method of using machine learning techniques to test for correlation between physical and operational characteristics of level crossings and observed collisions. The test identified that, overall, the *decision tree* method appears to be the most sensitive method for assessing whether any correlation exists.

Contribution 14: The test has identified that a small degree of correlation can be found between operational characteristics and collisions, however no meaningful correlation can be found with the data that are collected on hazards and controls. This finding does not necessarily imply that no correlation exists, rather that a correlation cannot be found with the available data using the methods that have been applied.

Chapter 8: Discussion and summary of contributions

This chapter considers the implications of the findings of this study and summarises the contributions of the study.

8.1 Implications of the findings of the study

An important finding of this study is that the rate of collisions per road user traverse is lower at level crossings where there are more road users. This effect has not been identified by other researchers. The focus of this study has been to examine the data on observed collisions and provide a description of the collision rates; it has not been to posit hypotheses for why the rates of collision might vary with road traffic volumes. Nevertheless this study has identified a pronounced effect that is worthy of further consideration. There are a number of reasons why this effect may be seen, including reasons that arise as a result of road traffic operations; railway operations; or human behaviours. These potential reasons are considered in turn below.

8.1.1 Road traffic operations

When considering road traffic operations, in general there is a correlation between the number of road users (V) traversing a level crossing and the class of level crossing: railway-controlled level crossings are generally found where V is higher, and passive level crossings are generally found where V is lower (Evans and Hughes, 2019). However the effect of the rate of collisions per road user traverse falling at higher values of V is seen *within* each class of level crossing. As such, this study has controlled for effects that may be caused by the different types of warning device associated with each class of level crossing.

A possible reason for the effect was posited by Stott (1987) who argued that the first road user to stop at a level crossing will cause an obstruction to following road users. By this argument it is only the first road user who has an opportunity to collide with a train. As road traffic volumes increase at a level crossing, a larger proportion of the road users waiting at a level crossing will be in a queue behind the first vehicle. In this way it should be expected that the rate of collision falls away with increased values of V.

This study has found that actual rates of collision do not correlate well with the numerical values predicted by Stott; however this finding does not mean that Stott's underlying reasoning is faulty. Rather it may be that Stott's reasoning was sound but the formulation of the equation was not sufficiently rigorous to describe real-world effects. For example, Stott's equation does not consider the fact that the first vehicle to stop at a level crossing may be a small vehicle, such as a motorcycle, that could easily be passed by subsequent road users; nor that if the second vehicle in a queue were a heavy vehicle that failed to stop, it could easily push a light vehicle into the level crossing. Despite the fact that Stott's formulation does not exactly match observation, the reasoning in his hypothesis is compelling and may be sufficient by itself to explain the observed effect.

Another cause related to road traffic operations that may explain the observed effect could be a result of the method used to count road traffic volumes. The data provided by Network Rail, and used in this study, are the counts of road users traversing a level crossing in a day. A simplistic assumption is that when there are more road users traversing a level crossing in a day, then there must be correspondingly more road users queued during each operation of the level crossing. Such an assumption may not be valid.

It is easy to imagine that the busiest level crossings occur in urban areas where there are both a large number of road users and train movements. In these urban areas, it is likely that is a larger number of adjoining roads than at level crossings in rural areas. Consequently there may be more alternative routes for road users to take to avoid a level crossing. When road users observe that the warning sequence has started, or a train is approaching, they may elect to take an alternative route to avoid the level crossing. The alternative route may be longer than the route over the crossing, but may be preferable to waiting. If this were the case, then the number of road users waiting at a level crossing on each train approach would not necessarily increase in proportion to the total number of road users traversing a level crossing in a day. If this were the case, then the rate of collisions per road user may not increase as expected with increased road traffic.

8.1.2 Railway operations

As noted above, level crossings with larger numbers of road users can be expected to occur in urban areas where there are also more train movements. The larger the number of train movements that occur at a level crossing, the more likely it is that two trains will traverse the level crossing at the same time. Whilst these simultaneous train movements are recorded as two separate traverses in the data provided by Network Rail, from the perspective of a road user the effect is the same as a single train traverse. Once a road user has stopped clear of a level crossing to wait for a train to pass, it does not matter whether one train or two pass during the time the level crossing is closed. In these cases, the rate of collisions per road user per train traverse will, in effect, be halved.

It should be noted that this argument may be true only in cases when the train traverses are actually simultaneous. When the traverses are not simultaneous but are closely spaced, the warning devices at a level crossing continue to operate after a train has traversed to close the road for another train that is approach shortly. In these cases the subsequent train may present an increased risk to road users, as road users may become impatient, or may believe that the warning equipment is faulty, and might move into the level crossing into the path of the subsequent train. In these cases the safety risk may increase as a result of increased rail traffic.

8.1.3 Human behaviours

Aside from the reasons noted above regarding road traffic operations, it is possible that road users change their behaviour in the presence of each other in a way that reduces the risk. There are some studies that have identified that compliance to rules increases when people believe they are being observed. The phenomenon is known as the *Hawthorne effect*, although the exact nature of the effect is disputed (Wickström and Bendix, 2000) and no studies can be found that investigated this effect at railway level crossings. However Barić *et al.* (2018) studied road user compliance at level crossings and noted that when a uniformed police officer was present "*the proportion of illegal crossings by pedestrians and cyclists alike fell nearly to zero*". It is possible that when there are more people around, road users assume that there is an increased likelihood of being penalised for breaching warnings. Such penalties may be the direct intervention of law enforcement, or may simply be the loss of an individual's self-esteem if they believe that others observing them are judging their behaviour. In either case the effect would be that the rate of non-compliance, and consequently the rate of collisions,per road user traverse can be expected to reduce when there are more road users.

Another possibility is that road users may have lower expectations of traffic speed at busier level crossings. Where there are more road users, individuals may be expecting to be delayed for some reason, including traffic congestion, and may be more willing to wait at a level crossing. Conversely in rural areas, road users may expect to be able to continue unimpeded on their journey; a level crossing may be seen as an unwelcome obstacle in their journey that may elicit lower rates of compliance, and consequently higher rates of collision, per road user.

8.2 Summary of contributions

Table 8.1 provides a summary of the contributions to current knowledge on level crossing safety have been established as a result of this work.

Table 8.1: Summary of contributions

Contribution 1:	It has been identified that there is a gap in the knowledge of SRPTs for level crossing and that there is an opportunity to advance the current state of knowledge by using newly available sources of data and by combining data sources on level crossings and observed collisions in a way that has not previously been performed.
Contribution 2:	This work has reviewed the literature that are available on level crossing safety and in particular SRPTs and the traffic models that underpin them. The review has identified the sources of data that are available for validating the SRPTs and methods of testing traffic models including emerging machine learning techniques.

Contribution 3:	Two different methods were used to show that traffic moment is a valid normaliser of collisions at level crossings in simple cases where there is: unsaturated traffic; a single carriageway approach in each direction; no queuing from nearby roads; and road users take no care to avoid a collision.
Contribution 4:	This study has established *collisions per train traverse per day* as a unit that provides a meaningful way to compare collision rates between level crossings.
Contribution 5:	A rigorous review was undertaken to identify a suitable test to allow a meaningful comparison of observed collision rates against traffic models. In particular the method of testing needs to be robust in cases where data are *overdispersed*. The *Kolmogorov Smirnov* test was identified as being appropriate for this analysis.
Contribution 6:	The study described undertook rigorous tests to compare observed collisions with proposed traffic models. The method was performed in a repeatable manner that would allow the test to be carried out with any other traffic model that may be proposed.
Contribution 7:	Whilst it was shown that traffic moment can be a normaliser for collisions in idealised conditions, in real-world conditions it does not appear that observed collisions vary in accordance with traffic moment. This finding may have profound implications for the many SRPTs that use traffic moment as an underlying traffic model.
Contribution 8:	Similarly, it does not appear that in general Stott's hypothesis is a meaningful normaliser for observed traffic collisions. Specifically Stott's hypothesis was developed to describe collisions at automatic vehicular level crossings, however the Peabody Dimmick model appears to correlate better with observation in this class of level crossing.

Contribution 9:	The descriptive traffic model developed by Peabody Dimmick is no worse at describing the rate of level crossing collisions than the predictive hypothesis developed by Stott. This finding is particularly noteworthy since the Peabody Dimmick model was developed approximately 90 years ago. This finding may indicate that the SRPT which uses Stott's hypothesis as a traffic model may produce more accurate results if the Peadbody Dimmick model were used instead.
Contribution 10:	In general, none of the proposed traffic models correlated with observed collisions with any meaningful degree of statistical significance.
Contribution 11:	The study identified that the distribution of collisions over road traffic volume (V) does, in every case, vary in accordance with a power law of the form *collisions per train traverse per day* $= a'V^{b'}$. In every case the value of b' is below zero.
Contribution 12:	The corollary of the observed power law indicates that level crossing closure can be considered an effective method to reduce the total number of collisions. This finding is particularly significant as the infrastructure manager of the GB railway has been undertaking a programme of level crossing closure although, to date, no studies have been undertaken to indicate that such an approach can be expected to improve safety overall.
Contribution 13:	The study has demonstrated a method of using machine learning techniques to test for correlation between physical and operational characteristics of level crossings and observed collisions. The test identified that, overall, the *decision tree* method appears to be the most sensitive method for assessing whether any correlation exists.
Contribution 14:	It was identified that a small degree of correlation can be found between operational characteristics and collisions, however no meaningful correlation can be found with the data that are collected on hazards and controls. This finding does not necessarily imply that no correlation exists, rather that the data that are collected do not allow any correlation to be found.

Each of these contributions is discussed below.

8.2.1 Contribution 1 – Knowledge gap regarding level crossing safety

The literature review in Chapter 2 led to a number of findings. Most importantly is that there is a large corpus of literature on level crossing safety although the primary finding is that whilst there are a large number of works that describe individual studies of potential causal mechanisms for collisions at level crossings, there is no generally agreed-upon theory on level crossings collisions, their underlying causes and the most efficient means to improve safety. Furthermore none of the other researchers in this area has previously highlighted this lack of an over-arching theory. It would not be unfair to state that the general state of research on level crossing safety is directionless.

The literature appears to agree on only two general themes firstly is that there is a general assumption that the physical and operational characteristics of a level crossing affect road users' situational awareness and motivation to stop at a level crossing, which in turn affects the likelihood of a collision (refer to Figure 2.2). It is therefore believed that interventions that affect either the physical or operational characteristics of a level crossing will necessarily affect the rate of collisions. Despite this belief being widespread within the literature, it is both tacit and unexplored. No prior research has put forward this simple statement regarding the belief that physical and operational characteristics indirectly affect the likelihood of a collision, nor has any controlled experimentation been performed to test the limits of any such causal influence.

The second general point of consensus is that there is a hierarchy of warning devices that can be applied at level crossings that correlate with reduced safety risk as the hierarchy is ascended. It is generally believed that passive warning devices correlate with greater safety risk that active devices; and barriers correlate with lower safety risk than open level crossings. This belief is not tacit as a number of researchers refer explicitly to the hierarchy, however whilst there is broad agreement regarding the form of the hierarchy there is no consensus regarding its exact form and the number of layers in the hierarchy. Furthermore non-physical controls, such as education and enforcement programmes which are valid controls for reducing level crossing safety risk, are never considered to be part of the hierarchy. The belief in the hierarchy has been long-held but it is only recently that it has been proven correct by rigorous study.

8.2.2 Contribution 2 – Proliferation of SRPTs and traffic models

The literature review in Chapter 2 identified that there are an abundance of SRPTs that are used to attempt to determine safety risk at level crossings around the world. Each of the models has an underlying traffic model which in most cases is traffic moment. There are a number of important observations regarding this proliferation of SRPTs.

The first observation is that the abundance of SRPTs is at odds with the lack of an over-arching theory of level crossing collision causation. It can reasonably be expected that SRPTs encode the knowledge of a well-test theory of level crossing safety. Yet without any such theory it has to be questioned how the tools have been determined. This absence of an underlying theory of level crossing safety is highlighted by the fact that SRPT used in Britain was fundamentally changed by switching its underlying traffic model (from *traffic moment* to *Stott's hypothesis*). It must be presumed that this change was necessary since the tool had previously not been producing accurate results. This is a serious consideration since the tools are used to direct spending to improve public safety, if the tools are not grounded in a robust theory, then it is not clear that the objective of public safety can be achieved. The second observation extends the first; from what is known of the various SRPTs it is clear that they do not agree with each other in their methods of risk calculation. Since it can be expected that the varying methods of calculation produced varied results, and it can also be expected that some tools produce more accurate results than other, it must be inferred that there are some SRPTs being used that may not be fit for purpose.

The third observation is that, in general, there is not good information regarding the method of calculation used in the SRPTs. For a small number of the tools there is a full description of the method of calculation, but these are the exception. For some tools there is no information and for the majority there is only sparse information: certainly not sufficient for the method of calculation to be reproduced and tested. The fourth observation extends the previous point: in the absence of transparency regarding the method of calculation, public confidence in the tools could be created if there were evidence that rigorous, independent testing of the tools had been undertaken to demonstrate the correctness of the tools and their fitness for purpose. Again, no such evidence can be found. The final observation is that there is an emerging science in data-driven safety risk management for the railways, however this approach does not appear to have been extended to level crossing SRPTs.

Overall the review of literature on SRPTs creates a bleak picture: there is no reason to believe that any of the tools provide accurate risk predictions yet railways are continuing to use them to direct spending on public safety programmes. There does not appear to be any work currently being undertaken to test the approach that is being used or to bring it up-to-date with modern approaches for railway safety risk management.

8.2.3 Contribution 3 – Validation of traffic moment

Following from the findings of the review of prior work, a simple study was undertaken to test the traffic moment as a meaningful normaliser of collisions under simple conditions. The study took two approaches, the first a mathematical derivation using a simple dynamic model of trains and road vehicles approaching a level crossing. The results were that the rate of collisions can be expected to be exactly proportional to traffic moment. The second study used a Monte Carlo simulation using 630 million random trials over a range of scenarios. The simulation showed a very high degree of correlation between collision rates and traffic moment. Whilst the tests of traffic moment were simple and the results are not particularly surprising, this work provides a rigorous argument that traffic moment can be used as a normaliser for level crossing collisions. This study therefore provides a contribution that has long been missing from this field of study.

Finally the study found that whilst there may be a correlation between collisions and traffic moment for simple cases, it cannot be expected that the correlation will hold under all conditions, or even under any real-world conditions, since complex effect such as oversaturation change road traffic behaviour in unpredictable ways.

8.2.4 Contribution 4 – Unit of level crossing safety

In order to test observed collision rates against the traffic models it is necessary to identify a test variable. Consistent with there being no generally agreed-upon theory of level crossing safety, there is no generally agreed-upon unit of level crossing safety. Instead there are a range of measures that have been used at different times. A contribution of this study is to identify a suitable unit of level crossing safety which can be used to compare not only observed collisions with traffic models but also in the more general analysis of level crossing safety for example to compare overall performance between different categories of level crossing. The unit determined is: *collisions per train traverse per day*.

8.2.5 Contribution 5 – Statistical methods to compare overdispersed data

A further contribution of this study is to identify an appropriate statistical test that can be used to test correlation where there are overdispersed, non-parametric data. A rigorous approach was taken during the review to consider all statistical tests identified in the literature and consider their applicability for this study. A number of tests which may *prima facie* appear to be applicable, such as the Mann Whitney U test and the Shapiro Wilk test were discounted during this review. The only test identified that it suitable for the test variable in this study is the *Kolmogorov Smirnov* test. This finding can be applied for future studies on level crossing safety risk.

8.2.6 Contribution 6 – Analysis of traffic models

A significant contribution of this study was to perform the test of correlation between the observed collisions at level crossings in Britain and the three traffic models for which the test could be performed: traffic moment, Stott's hypothesis, and Peabody Dimmick's model. For all except one of the 36 tests performed, the Kolmogorov Smirnov test suggested that the null hypothesis should be rejected: observed collisions cannot be considered to have been sampled from the same distribution as any of the traffic models. Only for one of the 36 tests did the test not suggest rejection of the null hypothesis and that was only at the most permissive value of $\alpha = 1\%$ for the Peabody Dimmick model when compared with collisions at automatic vehicular public level crossings using SMIS data. Further analysis of the test results, however, showed that none of the models correlates well with observation in all cases and the best correlating model varies depending on the category of level crossing.

Following on from the observations above regarding the proliferation of untested SRPTs and the lack of transparency in the method of calculation, the main observation to be made at this point is that the only test that can be made given the furtive nature of the SRPTs does not support the belief that the tools are generally fit for purpose. This finding is very serious given the use of the tools to direct spending on expensive level crossing warning devices for reasons of public safety.

An important outcome from this study is the degree to which the Peabody Dimmick model corresponds with observation, especially compared with other models. The Peabody Dimmick model is a descriptive model based on observation of collisions in Illinois in the 1930s. By contrast, Stott's hypothesis was created specifically to describe collision rates at automatic level crossings in Britain in the 1980s. Yet the Peabody Dimmick model outperforms Stott's hypothesis for the exact class of level crossings it was created to described. Since the 1930s there have been significant changes to technology: trains are no longer steam-powered; there have been changes to the design of road vehicles; a greater proportion of roads are paved. Furthermore there are more vehicles and road users have more exposure to traffic and therefore can be expected to have acquired different skills for driving. Of course, the largest difference is that the Peabody Dimmick model was derived from observed collision rates in Illinois, not Britain. It might appear remarkable that there appears to be any correlation at all. The implication is that there may be some underlying invariants that affect level crossing safety that are immutable across different continents and approximately 90 years. Again in the absence of a general theory of level crossing collision causation, it can only be speculated what these invariants may be.

8.2.7 Contribution 7 – Use of traffic moment as a normaliser

Whilst there is theoretical evidence to support use of traffic moment as a normaliser in idealised conditions, there is poor correlation between traffic moment and observed collisions. Traffic moment is the most commonly used traffic model amongst the SRPTs studied by previous researchers (RSSB, 2007). This finding could have a profound impact for these SRPTs. Since no evidence can be found of validation of any SRPT, it is not clear whether the tools produce accurate results, however the work that was performed in this study casts doubt on the traffic model that underlies these tools and, as such, gives rise to further concern about the suitability of these tools for determining level crossing safety interventions.

8.2.8 Contribution 8 – Use of Stott's hypothesis as a normaliser

The SRPT used in Britain (the *ALCRM*) was initially based on traffic moment, however a change was made to instead use Stott's hypothesis. This change "*caused a significant re-appraisal of which are the highest-risk level crossings in GB. Some crossings are now shown by the ALCRM to be relatively higher risk than previously thought, while other more busy crossings may actually be safer*" (Baker and Heavisides, 2007). However there is no evidence that the change led to risk predictions that were any more accurate than would previously have been made using traffic moment as a traffic model.

The work carried out in this study showed that, overall, Stott's hypothesis does appear to produce more accurate results than traffic moment, the accuracy of Stott's hypothesis is, in some cases, poor. Again, this finding raises a concern regarding the suitability of the ALCRM for determining safety interventions at level crossings.

8.2.9 Contribution 9 – Suitability of descriptive models

It was found that, in many cases, the Peabody Dimmick traffic model provided more accurate predictions of collision rates than the other traffic models. This finding is particularly salient since the Peabody Dimmick model is a *descriptive* model that was developed around 90 years ago based on observations in the United States. Conversely, Stott's hypothesis a *predictive* model that was developed in the 1980s specifically to describe level crossing collision rates in Britain. It is therefore remarkable that there are cases where the Peabody Dimmick model produces more accurate results than Stott's hypothesis for level crossing in Britain.

This finding suggests that there may be underlying behaviours of road users that affect rates of collisions and that remain relatively constant regardless of significant changes in road and railway technology. The corollary of this finding is that the most effective means to develop either traffic models or SRPTs, may be to develop descriptive tools based on observed road user behaviour.

8.2.10 Contribution 10 – Poor correlation of any traffic models

Overall the poor correlation of the traffic models tested casts doubt on the predictive accuracy of any of the SRPTs in use around the world. This study has identified that this finding is not necessarily the result of there being a small number of level crossing against which to test the models. In some cases the classes of level crossing with fewer test samples produced better correlation with traffic models than those with more samples. Rather it can be concluded that, in general, the traffic models, do not produce good predictions of collision rates.

It is noted that the traffic models are only one component of SRPTs, as such it may be that whilst the traffic models themselves do no produce accurate predictions, the SRPTs may produce good predictions by giving consideration to other factors such as the hazards and risk controls in place at level crossings.

8.2.11 Contribution 11 – Observed power distribution of collisions over V

Having determined that the extant traffic models do not correlate well with observed collisions, a further contribution of this study has been to identify, for each class of level crossing, a general distribution of how rates of collisions vary over road traffic volume. The study identified that, in general, the distribution of collision rates appears to vary in accordance with a simple power relationship of the form *CP(collision per train traverse per day)* $= a'V^{b'}$. In all cases the value of b' is negative indicating that, in general, higher road traffic volumes correlate with fewer collisions per road user traverse.

The major contribution from this analysis is that the relationship between road traffic volume and collisions can be described by a simple power equation which, in most cases, corresponds well with observed collisions. This is a significant contribution using a new source of information and analysis of collision data from the railways in Britain. The identification of the power function does not confirm a previous hypothesis posited by any other researcher, it is an entirely new finding based only the prior study described in this work.

The presence of the power relationship raises two questions; firstly why should any such relationship exist, especially since, in some cases, it appears to correlate strongly with observation? Unfortunately the absence of a general theory of level crossing safety prevents any explanation of the phenomenon, again the state of current knowledge renders any attempt at explanation to be nothing more than speculation. The second question is whether the relationship is generally true across all level crossings? The fact that the Peabody Dimmick model appears to correlate in some cases with observed collisions in modern Britain suggests that there may be underlying invariants in level crossing collision causation. In which case the identification of such a distribution within British level crossings may suggest that the same distribution can be found elsewhere around the world. At the current time, the data to perform such an analysis are not available.

The other observation arising from this work might be considered to be a validation of a pre-existing hypothesis, but is nevertheless highly significant: that level crossing closure can be undertaken as part of a meaningful programme to reduce the overall number of collisions at level crossings. The railway infrastructure manager in Britain has a policy to close level crossings wherever possible. Until now there has been no evidence that such a policy could be expected to reduce the overall number of collisions at level crossings. Hypothetically, if the rate of collisions per traverse were found to increase as the volume of level crossing users increased (if b' were found to be positive), the closing level crossings would have an adverse effect on level crossing safety. The discovery from this work that b' is negative for level crossings in Britain (and potentially elsewhere) is a significant contribution to the practical management of level crossing safety.

8.2.12 Contribution 12 – Level crossing closure as a meaningful intervention

The railway infrastructure manager in Britain has undertaken a programme of closing level crossings wherever possible as a means of improving safety. However there is no prior evidence that level crossing closure is an effective means to improve safety overall. However when a level crossing is closed, it cannot be assumed that road users will abandon any desire to travel to the other side of the rail, rather road users will traverse the rail at other places, including other level crossings. It is not clear that causing road users to divert from one level crossing to another will necessarily reduce the overall number of collisions; in fact it could potentially lead to more collisions. This study has shown that the number of collisions *per road user traverse* decreases as the volume of road users increases at a level crossing. This finding provides an important validation of the programme of level crossing closure as a means to improve safety overall.

8.2.13 Contribution 13 – Use of machine learning instead of statistical methods

Determining a relationship between road traffic volume and collision rates is a contribution of this study. In effect this work has developed descriptive traffic models for twelve classes of vehicular level crossings in Britain. Since these models have been derived based on empirical data they could be used in place of the predictive models that have previously been used in the ALCRM (being traffic moment and Stott's hypothesis). There are two fundamental difficulties in attempting to create any descriptive model. The first lies in being able to determine which variables to include in the model. The reason SRPTs are created in the first place is to allocate resources to provide safety risk controls at level crossings. It is not expected that the safety risk at level crossings can ever be reduced to zero, rather it is expected that provision of suitable controls can reduce the risk to an acceptable level. As such there is an implicit admission that the exact causes of collisions can never be entirely known nor controlled. After all if it were known in advance exactly what would be the cause of a specific collision then the railway would take necessary action to prevent the collision occurring. Instead the approach is to take all reasonably practicable actions to reduce the risk within the constraints of the resources available. As such, since it can never be clear exactly what the causes of collisions are, there is no way to be certain that all necessary variables have been included in an SRPT. Conversely, creation of a predictive model assumes that it is possible to establish in advance what variables are necessary – and by extension – which are not necessary to determine safety risk.

The second difficulty lies in being able to determine how to combine the variables to create a model with good predictive accuracy. For example if, in constructing a predictive model, it were believed that the speed of trains contributed to safety risk the question would be in what way does speed affect the risk? Is the relationship linear: does doubling a train’s speed double the safety risk? Or is there some other relationship? Similarly there is a question regarding the relative contributions of different factors: does sun glare contribute the same amount of risk as train speed, or are the relative contributions different by some factor, if so what factor? A contribution of this work is that the tests carried out have established that the predictive models available do not outperform a descriptive model.

Depending on other factors that are used to calculate risk scores in the tool, it can reasonably be expected that use of more accurate traffic models will produce improved risk predictions than are currently produced by the ALCRM. It may be tempting to subject the derived traffic models to a predictive accuracy test using the approach described in Chapter 6, however any such test would be tautological and meaningless: the test would be performed using the same data that were used to derive the models. Whether the models would continue to perform well into the future is an important test that could be performed at some point in the future.

It is likely that the traffic models derived in this study could be improved by using a richer source of data than is currently collected to calculate risk predictions from SRPTs. Considering the model described in Figure 2.2 which shows the presumed relationship that the physical and operational characteristics of a level crossing affect road users' situational awareness and motivation to stop which, in turn, affect the likelihood of a collision. It is notable that the current methods of study in level crossing safety explicitly collect data on the first and third elements in this chain of causation, *viz.* physical and operational characteristics; and occurrence of collisions. However no data is explicitly collected on the middle element in the chain: road users' situational awareness and motivation to stop. It is possible that the lack of explicit consideration of road users' behaviour is a reason why it is not possible to find a meaningful correlation between the recorded characteristics of level crossings and observed collisions. Furthermore it is possible that for this reason the SRPTs in use around the world do not provide accurate risk predictions, which may explain why no evidence of validation of the models is available. Collection of such data would allow for the middle stage of Figure 2.2 to be considered as part of the method of calculation in an SRPT.

It is understood that explicit collection of data regarding road users' situational awareness and motivation to stop is difficult as these are properties of the road users' minds and currently there is no technology that allows road users' thoughts to be readily collected. However it is possible that proxy data could be collected which may allow reasonable inferences to be made about road users' states of mind at a given time; for example, data could be collected from level crossing obstacle detectors. These data can provide an insight into the occasions when road users approach a level crossing and choose not to stop. Other technologies, such as video cameras, may be employed to detect how road users change speed on the approach to a level crossing. It is likely that there are observable patterns of behaviour that correspond with different levels of situational awareness. For example there may be specific changes in vehicle speed when a vehicular road user only becomes aware of the need to stop at a level crossing a few seconds before arriving at the level crossing. Alternatively there may be different types of changes in road speed that occur when road users approaches a level crossing with caution, then purposefully decides to breach the warning since they believe they can traverse safely before the arrival of a train.

As well as possibly allowing for inference of road users' situational awareness, collection of data from obstacle detectors would also have the advantage of inflating the available data when considering collisions at level crossings as obstacle detectors would allow collection of data on *near-collisions*, for example instances where a road user is occupying a level crossing until only five seconds before the arrival of a train. The larger source of data would reduce the likelihood of the derived models over-fitting the observation data and therefore improving the general predictive accuracy of the models. However it is necessary to raise a caveat: if near-collision data were to be used there are two important points to consider: firstly is that not all level crossings are currently fitted with obstacle detectors. It would therefore be necessary to determine an appropriate method of calculation so that the absence of near-collision data is not interpreted as an absence of near-collisions. The second consideration is that near-collisions are not collisions. In may be that in some cases the causal factors that lead to near-collisions are exactly the same as those that lead to collisions, in which case the difference between collisions and near-collisions is merely a matter of uncontrollable random chance and near-collision data can be used as a proxy for collision data. In other cases there may be significant differences between the causal factors that lead to collisions and those that lead to near collisions. In such cases the use of near-collision data would be inappropriate. Again the absence of a general theory of level crossing collision causation means that currently it is not possible to determine whether near-collisions data can be used in this way. Where near-collision data are made available, then a future study should be undertaken to determine the degree to which near-collisions correlate with collisions and therefore the degree to which they can be used as proxy data.

It is also observed that if near-collision data can be used in this way, then it would be clear that obstacle detectors can have some use to the railway. Currently, the long braking time of trains and the need to run an efficient railway mean that it is not clear in what way obstacle detectors can be used to the benefit of railway safety.

A final observation is that the use of Stott's traffic model to underpin the ALCRM when there is no transparency in the method of calculation and no evidence of accuracy is an undesirable situation for public safety and should not be tolerated by policymakers. This study has demonstrated how an improved traffic model can be derived using data that are currently available. The emergence of additional sources of near-collision data and new approaches to data-driven safety management together with machine learning techniques provide all the components that are needed to create an accurate and transparent SRPT for the railways in Britain. The new data and emerging technologies also provide for the opportunity to more fully understand level crossing collisions that may allow for the creation of a consistent and complete theory of level crossing safety that has eluded the field of research until now.

8.2.14 Contribution 14 – Poor correlation of hazards and control data with observed collisions

Using the available data on level crossings, together with advanced machine learning methods, a study was undertaken to determine whether the occurrence of hazards and risk controls at level crossings correlated with observed collisions. The study found that whilst road and rail traffic volumes correlate with collisions, no other correlations can be found. This finding reinforces the earlier findings in this work that there is some degree of correlation between traffic models and collision rates. However no meaningful correlation could be found between collisions and the data on hazards and risk controls. *Per se*, this finding may not indicate that hazards and risk controls do not affect rates of collision, rather the finding may indicate that the data collected on level crossings do not contain the correct items for a correlation to be found. It is possible that if data were collected regarding other features of level crossings then meaningful correlations could be found.

Another possibility is that there really is no correlation. The finding in Contribution 9 indicates that there may be human factors that affect road user behaviour at level crossing and that these factors are largely invariant regardless of 90 years' of advances in road and rail technology. It is further possible that the factors that affect the likelihood of collisions at level crossing are not affected by the physical features of a level crossing, rather it may be that the fundamental behaviours of road users have the largest impact.

Chapter 9: Conclusions

This chapter summarises the findings of this study and discusses further work that could be undertaken to address the findings of this study.

9.1 Summary of findings

There is widespread use of SRPTs around the world, however the lack of evidence of the accuracy of these tools is a concern for public safety. The work in this study casts doubt on the accuracy of the traffic models that underpin these tools and on whether the presence of hazards and risk controls at level crossing have any meaningful correlation with collision rates. It is possible that, in principle, an SRPT could be developed that has good predictive accuracy, however it is not clear that the data that are currently collected allow such a tool to be created. Rather, this study indicates that there may be underlying invariants in human behaviour that affect the likelihood of a collision at a level crossing.

All SRPTs are developed on the assumption that the physical and operational characteristics of a level crossing affect road users' situational awareness and motivation to stop at a level crossing, which in turn affects the likelihood of a collision as described in Figure 2.2. However none of the SRPTs explicitly address the question of road users' situational awareness and motivation to stop. Rather it appears to be assumed that if sufficient data are collected on the physical and operational characteristics of level crossings then an accurate SRPT could be created nonetheless: this study finds that this assumption does not appear to be valid. The number of variables that affect level crossing safety may be so large that the task of determining how to combine the data to produce a meaningful risk prediction is intractable. Consideration needs to be given to developing an SRPT that explicitly considers human factors in determining safety risk.

The concept of using SRPTs to determine safety interventions at level crossings is not fundamentally faulty, however it seems that the application of SRPTs to date may not have been effective. It is possible that in the future an accurate tool could be developed. Such a tool should be based on a descriptive model determined from real-world observation, and would require a much broader range of data than has been collected to date. Advanced data analysis techniques such as machine learning can be used to support the development of such a tool. As well as the use of emerging technology to develop an appropriate model, it is possible that other new technologies can be applied to overcome some of the other difficulties in determining level crossing safety. For example obstacle detectors could be used to collect data on near-collisions; such extra data would overcome some of the problems encountered during this study due to the difficulties in analysing overdispersed data.

From the evidence considered in this study, it appears that the approach of closing level crossings could be effective in reducing the overall number of collisions. This outcome can be expected regardless of the traffic volume at the level crossing; in particular a reduction in the total number of collisions can be expected even when the level crossing being closed has only a small traffic volume. This finding supports the approach being undertaken by the railway infrastructure manager in Britain to close level crossings wherever possible bearing in mind the broader needs of road users and the community.

Through a rigorous approach of using the best available data and emerging machine learning methods, it has been identified that the means are available to produce a safety risk prediction tool that can be expected to be more accurate than the ones currently in use. It is hoped that the railway act on the results obtained from this study in the very real interests of public safety.

9.2 Further study

This study has identified some areas where there is a need to address shortcomings in the current knowledge of level crossing safety. Possible areas for further study are considered below.

9.2.1 Development of an over-arching theory of level crossing safety

There has been a significant amount of work carried out by a number of researchers investigating level crossing safety. Whilst individual studies have produced meaningful results, there is no over-arching theory of level crossing safety within which the results of individual studies can be understood. Currently the results of various researchers may complement or contradict each other, however there is no structured theory for understanding where such overlaps in understanding occur. Furthermore, in the absence of a general theory, there is no clear way to identify where there are gaps in the current understanding and to direct future research. It can be expected that a general theory of level crossing safety would inform the development of an SRPT that correctly addresses the factors that affect safety at level crossings, as well as providing the means to allow validation of such a tool.

9.2.2 Study of road user motivation and situational awareness

This study has highlighted the need to undertake further study of road users' situational awareness and motivation to yield to trains at level crossings. It is proposed that a study would start by creating a formal model of road user interactions with each class of level crossing, which would detail:

- what behaviour is required from road users including the required behaviours on the approach to the level crossing as well as during the traverse;
- what information is provided to road users in the form of signs, lights, audible warnings etc.;
- how failures of a road user to understand or comply with warnings can affect the likelihood of a collision;
- the critical points where incorrect behaviour can result in collisions and the factors that may cause errors or violations to occur.

In creating a formal model consideration should be given to the different actions required of a road user such as:

- yielding to an approaching train or a train that is already occupying the level crossing;
- ensuring that there is sufficient space to exit the level crossing after entering, and not forming a queue over the level crossing;
- being aware that a second train may be approaching a level crossing after the first train has cleared; and
- being aware of other hazards, such as the risk of grounding low vehicles, or contact with overhead electrical conductors.

The study should consider the various means of providing visual and audible cues to road users. To be generally useful, the study would need to consider how road users interpret the various cues, which is likely to involve some combination of psychological experimentation, interviews, focus groups, or direct observation of road user behaviour.

The model could consider how failure may occur in providing information to road users – for example by warning lights being obscured by vegetation or traffic – and whether secondary sources of information are available. The study should also explicitly consider how different enforcement actions affect road user behaviour.

A study of this nature has been proposed to Network Rail, and work is expected to begin within the next year.

References

ABC Railway Guide (2019). Your Guide to Britain's Railway Infrastructure. Retrieved from http://abcrailwayguide.uk/; 26 June 2019.

Aeron, S., Bose, S., Valero, H. P., and Saligrama, V. (2011). Broadband dispersion extraction using simultaneous sparse penalization. *IEEE Transactions on Signal Processing*, 59(10), 4821-4837. DOI: 10.1109/TSP.2011.2160632

Aggarwal, C. (2014). Data Classification: Algorithms and Applications, ser. *Frontiers in physics*. Chapman and Hall/CRC.

ALCAM (Australian Level Crossing Assessment Model) (2007) *Australian Level Crossing Assessment Model Technical Manual*; 10 February 2007

Alpaydin, E. (2009). *Introduction to machine learning*. MIT press.

Amari, S. I., Murata, N., Muller, K. R., Finke, M., and Yang, H. H. (1997). Asymptotic statistical theory of overtraining and cross-validation. *IEEE Transactions on Neural Networks*, 8(5), 985-996. DOI: 10.1109/72.623200

Amato, C., and Shani, G. (2010, May). High-level reinforcement learning in strategy games. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1* (pp. 75-82). International Foundation for Autonomous Agents and Multiagent Systems. DOI:10.1145/1838206.1838217

Antoneli, F., Passos, F. M., Lopes, L. R., and Briones, M. R. (2018). A Kolmogorov-Smirnov test for the molecular clock based on Bayesian ensembles of phylogenies. *PloS one*, 13(1), e0190826. https://doi.org/10.1371/journal.pone.0193633

Australian Transport Safety Bureau (ATSB) (2001); *Fatal crashes at level crossings*; September 2001.

Awad, M, Khanna, R. (2015). *Efficient learning machines: theories, concepts, and applications for engineers and system designers* (p. 268). Springer Nature.

Ayyadevara, V. (2018). *Pro Machine Learning Algorithms*. Apress.

Baker J and Heavisides J. (2007) *Traffic moment at level crossings: when busy crossings can be safer*. International Railway Safety Conference.

Barić, D., Pilko, H., and Starčević, M. (2018). Introducing experiment in pedestrian behaviour and risk perception study at urban level crossing. *International journal of injury control and safety promotion*, *25*(1), 102-112. https://doi.org/10.1080/17457300.2017.1341934

Beg, S., and Hasnain, M. S. (Eds.). (2019). *Pharmaceutical quality by design: principles and applications*. Academic Press.

Berk, R. A. (2008). *Statistical learning from a regression perspective* (Vol. 14). New York: Springer.

Binmore, K. (2008). *Rational decisions*. Princeton University Press.

Bonaccorso, G. *Machine Learning Algorithms: A reference guide to popular algorithms for data science and machine learning*. Packt Publishing, 2017.

Bzdok, D., Altman, N., and Krzywinski, M. (2018). *Points of significance: statistics versus machine learning*. https://doi.org/10.1038/nmeth.4642

Campbell, C., and Ying, Y. (2011). Learning with support vector machines. *Synthesis lectures on artificial intelligence and machine learning*, 5(1), 1-95. https://doi.org/10.2200/S00324ED1V01Y201102AIM010

Charan, J. (2010). How to select appropriate statistical test. *Journal of Pharmaceutical Negative Results* October, 1(2), 61. DOI:10.4103/0976-9234.75708

Crawford, W., and Csomay, E. (2015). *Doing corpus linguistics*. Routledge.

Dean, C. B., and Lundy, E. R. (2014). Overdispersion. *Wiley StatsRef: Statistics Reference Online*, 1-9. https://doi.org/10.1002/9781118445112.stat06788.pub2

DfT (Department for Transport, UK), (2019). *Statistical Release, Vehicle Licensing Statistics: 2019 Quarter 1 (Jan - Mar)*. 27 June 2019. Retrieved from: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data /file/830795/vehicle-licensing-statistics-april-to-june-2019.pdf. 26 August 2019.

Domingues, R., Filippone, M., Michiardi, P., and Zouaoui, J. (2018). A comparative evaluation of outlier detection algorithms: Experiments and analyses. *Pattern Recognition*, 74, 406-421. https://doi.org/10.1016/j.patcog.2017.09.037

Educba (n.d.). *Machine learning algorithms*. Retrieved from https://www.educba.com/machine-learning-algorithms/ 21 November 2019.

Evans, A. W. (2011). Fatal accidents at railway level crossings in Great Britain 1946–2009. *Accident Analysis and Prevention*, 43(5), 1837-1845. DOI:10.1016/j.aap.2011.04.019

Evans, A. W., and Hughes, P. (2019). Traverses, delays and fatalities at railway level crossings in Great Britain. *Accident Analysis and Prevention*, *129*, 66-75. DOI: 10.1016/j.aap.2019.05.006

Fakhfakh, N., Khoudour, L., El-Koursi, E. M., Bruyelle, J. L., Dufaux, A., and Jacot, J. (2010, July). Background subtraction and 3D localization of moving and stationary obstacles at level crossings. In 2010 *2nd International Conference on Image Processing Theory, Tools and Applications* (pp. 72-78). IEEE. DOI: 10.1109/IPTA.2010.5586765

Fausett, L. (1994). *Fundamentals of neural networks: architectures, algorithms, and applications*. Prentice-Hall, Inc.

Feigelson, E., and Babu, G. J. (2013). *Beware the Kolmogorov-Smirnov test*. Retrieved from: https://asaip.psu.edu/articles/beware-the-kolmogorov-smirnov-test/. 21 July 2019.

Gan, C. L. (2020). *Prognostics and health management of electronics: Fundamentals, machine learning, and the internet of things*. Wiley & Sons.

Garson, G. D. (1998). *Neural networks: An introductory guide for social scientists*. Sage.

Global Railway Review (2018), Network Rail: Improving level crossing safety with a multi-faceted approach, 05 June 2018; retrieved from https://www.globalrailwayreview.com/article/69816/network-rail-level-crossing-safety/ 12 March 2019.

Golio, M. (2015). Fifty Years of Moore's Law – Scanning Our Past. *Proceedings of the IEEE*, 103(10), 1932-1937. DOI: 10.1109/JPROC.2015.2473896

Google Trends (2019). Google Trends Explore. Retrieved from https://trends.google.com/trends/explore?date=all&q=machine%20learning 21 November 2019.

Grus, J. (2019). *Data science from scratch: first principles with python*. O'Reilly Media.

Gunawardena, N. (2011). Choosing the correct statistical test in research. *Sri Lanka Journal of Child Health*, 40(4). http://dx.doi.org/10.4038/sljch.v40i4.3838

Guo, Z., and Uhrig, R. E. (1992, June). Using genetic algorithms to select inputs for neural networks. In *[Proceedings] COGANN-92: International Workshop on Combinations of Genetic Algorithms and Neural Networks* (pp. 223-234). IEEE. DOI: 10.1109/COGANN.1992.273937.

Guyon, I. (1997). A scaling law for the validation-set training-set size ratio. AT&T Bell Laboratories, 1(11). Corpus ID: 16194090

Haleem, K. (2016). Investigating risk factors of traffic casualties at private highway-railroad grade crossings in the United States. *Accident Analysis and Prevention*, 95, 274-283. DOI: 10.1016/j.aap.2016.07.024

Hanea, D., Hanea, A., Ale, B., Sillem, S., Lin, P. H., Van Gulijk, C., and Hudson, P. (2012). *Using dynamic Bayesian networks to implement feedback in a management risk model for the oil industry*. In: 11th International Probabilistic Safety Assessment and Management Conference and the Annual European Safety and Reliability Conference 2012 (PSAM11 ESREL 2012). IAPSAM & ESRA, pp. 691-700.

Haykin, S. (1994). *Neural networks: a comprehensive foundation*. Prentice Hall PTR.

Hinton, P. R., McMurray, I., and Brownlow, C. (2014). *SPSS explained*. Routledge.

Hisamitsu, Y., Sekimoto, K., Nagata, K., UEHARA, M., and OTA, E. (2008). 3-D laser radar level crossing obstacle detection system. *IHI Engineering Review*, 41(2), 51-57. Retrieved from: https://www.ihi.co.jp/var/ezwebin_site/storage/original/application/618e8a5d0a842151631f835900e69615.pdf. 23 July 2019

Hollnagel, E. (2018). *Safety-I and safety-II: the past and future of safety management*. CRC press.

Hughes, P. (2002), A risk assessment system for passive level crossings, *Seventh International Symposium of Railroad-Highway Grade Crossing Research and Safety*, 20-21 February 2002, Melbourne, Australia

Hughes, P. (2012a), *Booms go bust - sensible safety at level crossings*, Institution of railway signalling engineers, Aspect 2012, 12 September 2012, London.

Hughes, P. (2012b). Booms go bust. *International Railway Journal*, *52*(12). Simmons-Boardman Publishing Corporation.

Hughes, P. (2014), WLX wireless level crossing warning system case study, *Rail Safety* 2014, Sydney, Australia. Retrieved from: https://www.informa.com.au/insight/the-next-generation-of-railway-level-crossing-safety-technology/. 15 February 2019

Hughes, P., Figueres-Esteban, M., and Van Gulijk, C. (2016, July). Learning from text-based close call data. In *Safety and Reliability* (Vol. 36, No. 3, pp. 184-198). Taylor & Francis. https://doi.org/10.1080/09617353.2016.1252083

Hughes, P., Robinson, R., Figueres-Esteban, M., and Van Gulijk, C. (2019). Extracting safety information from multi-lingual accident reports using an ontology-based approach. Safety science, 118, 288-297. https://doi.org/10.1016/j.ssci.2019.05.029

International Organization for Standardization. (2018). *Risk Management* (31000:2018).

Jordan, S., Mitterhofer, H., and Jørgensen, L. (2018). The interdiscursive appeal of risk matrices: Collective symbols, flexibility normalism and the interplay of 'risk'and 'uncertainty'. *Accounting, Organizations and Society*, 67, 34-55. https://doi.org/10.1016/j.aos.2016.04.003

Kanji, G. K. (2006). *100 statistical tests*. Sage.

Kelleher, A., and Kelleher, A. (2019). *Machine learning in production: developing and optimizing data science workflows and applications*. Addison-Wesley Professional.

Kyan, M., Muneesawang, P., Jarrah, K., and Guan, L. (2014). *Unsupervised Learning: A Dynamic Approach*. John Wiley & Sons.

Lam, H. K., Ling, S. H., Ling, S. S., and Nguyen, H. T.. (2012). *Computational intelligence and its applications: evolutionary computation, fuzzy logic, neural network and support vector machine techniques*. World Scientific. https://doi.org/10.1142/p773

Lantz, B. (2013). *Machine learning with R*. Packt publishing ltd.

Larose, D. T., Larose, C. D. (2014). *Discovering knowledge in data: an introduction to data mining* (Vol. 4). John Wiley & Sons.

Larue, G. (2016); *Waiting Times at Level Crossings Leading to Motorists' Risky Behaviours*; Australasian Centre For Rail Innovation. Retrieved from: https://www.acri.net.au/waiting-times-at-level-crossings-leading-to-motorists-risky-behaviours/. 15 February 2019

Li, F., Zhang, L., and Zhang, Z. (2017). *Dynamic Fuzzy Machine Learning*. Walter de Gruyter GmbH & Co KG. https://doi.org/10.1515/9783110520651

Mackridge, A., and Rowe, P. (2018). *A practical approach to using statistics in health research: from planning to reporting*. John Wiley & Sons.

Massey, F.J. Jr (1951) The Kolmogorov–Smirnov test for goodness of fit, *Journal of the American Statistical Association*, 4(6): 1990. https://doi.org/10.2307/2280095

McCrum-Gardner, E. (2008). Which is the correct statistical test to use?. *British Journal of Oral and Maxillofacial Surgery*, 46(1), 38-41. https://doi.org/10.1016/j.bjoms.2007.09.002

McElduff, F., Cortina-Borja, M., Chan, S. K., and Wade, A. (2010). When t-tests or Wilcoxon-Mann-Whitney tests won't do. *Advances in physiology education*, *34*(3), 128-133. DOI: 10.1152/advan.00017.2010

Mohammed. M, Khan. M B, Bashier. E B M,.(2016) *Machine Learning: Algorithms and Applications* 1st ed. CRC Press.

Molnar, C. (2018). *Interpretable machine learning: A guide for making black box models explainable*. E-book at https://christophm. github. io/interpretable-ml-book/, version dated, 10. Retrieved on 06 January 2020. Retrieved from: https://christophm.github.io/interpretable-ml-book/. 18 August 2019

MTCA Ministry of Transport and Civil Aviation (1957). *Level Crossing Protection*. A report by officers of the Ministry of Transport and Civil Aviation and of the British Transport Commission. HMSO, London. Retrieved from: https://www.railwaysarchive.co.uk/documents/MoT_LevelCrossing1957.pdf. 12 July 2019

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Nazari, Z., Yu, S. M., Kang, D., and Kawachi, Y. (2018, June). Comparative Study of Outlier Detection Algorithms for Machine Learning. In *Proceedings of the 2018 2nd International Conference on Deep Learning Technologies* (pp. 47-51). https://doi.org/10.1145/3234804.3234817

Network Rail (2015). *Transforming level crossings, a vision-led long-term strategy to improve safety on Great Britain's railway*.

Network Rail (2017). *Level Crossings Correct to Monday 28 December 2016*. Retrieved from http://www.networkrail.co.uk/transparency/level-crossings/ 15 March 2017

Network Rail, (2018). Annual report and accounts 2018, Network Rail.

Office of Rail Regulation (2011) Level Crossings: *A guide for managers, designers and operators* Railway Safety Publication 7, December 2011. Retrieved from: https://www.orr.gov.uk/media/10713. 18 February 2019

Oh, J., Simon P. Washington, and Doohee N, (2006) Accident prediction model for railway-highway interfaces. *Accident Analysis and Prevention* 38.2: 346-356. https://doi.org/10.1016/j.aap.2005.10.004

Ohta, M. (2005). Level crossings obstacle detection system using stereo cameras. *Quarterly Report of RTRI*, 46(2), 110-117. https://doi.org/10.2219/rtriqr.46.110

Oliveira, A. G., and Oliveira, A. (2013). *Biostatistics decoded*. Hoboken, NJ: Wiley.

Patience, G. S. (2017). *Experimental methods and instrumentation for chemical engineers*. Elsevier. https://doi.org/10.1016/C2010-0-66690-9

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., and Vanderplas, J. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.

Provost, F., and Fawcett, T. (2013). *Data Science for Business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media, Inc.

Quinlan, J. R. (1986). *Induction of Decision Trees*. Mach. Learn. 1, 1 (Mar. 1986), 81–106. https://doi.org/10.1007/BF00116251

Raschka, S. (2015). *Python machine learning*. Packt publishing ltd.

Rashidy, R. A. H. E., Hughes, P., Figueres-Esteban, M., Harrison, C., and Van Gulijk, C. (2018). A big data modeling approach with graph databases for SPAD risk. *Safety science*, 110, 75-79. https://doi.org/10.1016/j.ssci.2017.11.019

Renter, D. G., Higgins, J. J., and Sargeant, J. M. (2000). Performance of the exact and chi-square tests on sparse contingency tables. New Prairie Press New Prairie Press. In proceedings Conference on Applied Statistics in Agriculture 2000 - 12th Annual Conference. Retrieved from https://newprairiepress.org/cgi/viewcontent.cgi?referer=https://www.google.com/&httpsredir=1&article=1253&context=agstatconference. 18 August 2019.

RSSB (Rail Safety and Standards Board) (2007). Rail Safety and Standards Board. Use of risk models and risk assessments for level crossings by other railways. RSSB's Research and Development. *Rail Safety and Standards Board*. Report No, 21138.

RSSB (Rail Safety and Standards Board); (2018) *Signals, Handsignals, Indicators and Signs*; Handbook RS521; Issue 5; December 2018.

RSSB (Rail Safety and Standards Board); (2019) *Level crossing digest*; Version 1, January 2019 edition

RSSB. (2016) Rail Safety and Standards Board. Annual Safety Performance Report; A reference guide to safety trends on GB railways, 2015/16.

Russell, S. J., and Norvig, P. (2016). *Artificial intelligence: a modern approach*. Malaysia; Pearson Education Limited.

Sarker, S. D., and Nahar, L. (Eds.). (2018). *Computational Phytochemistry*. Elsevier.

Skovlund, E., and Fenstad, G. U. (2001). Should we always choose a nonparametric test when comparing two apparently nonnormal distributions?. *Journal of clinical epidemiology*, 54(1), 86-92. DOI: 10.1016/s0895-4356(00)00264-x

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., and Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, *15*(1), 1929-1958. DOI: /10.5555/2627435.2670313

Stamp, M. (2017). *Introduction to machine learning with applications in information security*. Chapman and Hall/CRC.

Starčević, M., Barić, D., and Pilko, H. (2016). Survey-based Impact of Influencing Parameters on Level Crossings Safety. *Promet-Traffic & Transportation*, 28(6), 639-649. https://doi.org/10.7307/ptt.v28i6.2208

Stevens, S. S. (1946). Stevens, S. S. (1946). On the theory of scales of measurement, *Science*, New Series, Vol. 103, No. 2684 (Jun. 7, 1946), pp. 677-680, American Association for the Advancement of Science. DOI: 10.1126/science.103.2684.677

Towards Data Science (2020). The 80/20 Split Intuition and an Alternative Split Method. Retrieved from https://towardsdatascience.com/finally-why-we-use-an-80-20-split-for-training-and-test-data-plus-an-alternative-method-oh-yes-edc77e96295d; 09 April 2021.

Transportation Research Board, National Research Council (2000); *Highway Capacity Manual*.

US DoT (United States Department of Department of Transportation), (2007). *Railroad-Highway Grade Crossing Handbook*. (No. FHWA-SA-07-010; NTIS-PB2007106220). Revised second edition.

Van Gulijk, C., and McCullogh, P. (2019, January). BowTies for the Digital transformation safety management systems. In *Hazards*. In proceedings 29th Institution of Chemical Engineers Symposium on Hazards 2019.

Van Gulijk, C., Hughes, P., and Figueres-Esteban, M. (2016). The potential of ontologies for safety and risk analysis. *Risk, Reliability and Safety: Innovating Theory and Practice*. London: Taylor & Francis Group, 1315-1322. Risk, Reliability and Safety: Innovating Theory and Practice – Walls, Revie & Bedford (Eds). 2017 Taylor & Francis Group, London.

Van Gulijk, C., Hughes, P., Figueres-Esteban, M., El-Rashidy, R., and Bearfield, G. (2018). The case for IT transformation and big data for safety risk management on the GB railways. *Proceedings of the Institution of Mechanical Engineers, Part O: Journal of Risk and Reliability*, 232(2), 151-163. https://doi.org/10.1177/1748006X17728210

Whitehorn, M. (2017). *Neural networks: Today, classifying flowers... tomorrow, Skynet maybe*. Retrieved from https://www.theregister.co.uk/2017/11/27/inside_neutral_nets/ 17 December 2019.

Wickham, M. (2018). *Practical Java Machine Learning: Projects with Google Cloud Platform and Amazon Web Services*. Apress.

Wickström, G., and Bendix, T. (2000). The "Hawthorne effect"—what did the original Hawthorne studies actually show? *Scandinavian journal of work, environment & health*, 363-367.

Wilde, G. J. (2014). Target Risk 3. *Risk Homeostasis in Everyday Life. Complimentary copy*, Web-version, 03-20. Retrieved from http://www.academia.edu/download/33747953/www16-2_TR3_working copy March 20 2014.docx; 12 May 2020

Witten, I. H., Frank, E. (2002). *Data mining: practical machine learning tools and techniques with Java implementations*. Acm Sigmod Record, 31(1), 76-77. https://doi.org/10.1145/507338.507355

Wujek, B. (2017). *Understanding Machine Learning Algorithms* Retrieved from https://www.kdnuggets.com/2017/10/understanding-machine-learning-algorithms.html 17 December 2019.

Wullems, C., Hughes, P., and Nikandros, G. (2013). Modelling risk at low-exposure railway level crossings: supporting an argument for low-cost level crossing warning devices with lower levels of safety integrity. *Proceedings of the Institution of Mechanical Engineers, Part F: Journal of rail and rapid transit, 227(5), 560-569.* https://doi.org/10.1177/0954409713501814

Appendix A: Collision data used in study

This appendix provides a list of the collision data supplied by RSSB and used in the study described in Chapter 5.

Table A.1: Collision data provided by RSSB

Event ID	Event date	Crossing type	Level crossing name	Territory description	Location description	ELR code
483282	11/04/1996	UWC+MWL		London North Eastern	Scorborough LC	HBS
505728	01/05/1996	UWC+T		London North Eastern	South Elmsall	DOL1
702902	20/06/1996	AHB		South East	Chartham Hatch LC	FDM0
689508	15/11/1996	AHB		London North Western	Bescar Lane LC	WBS3
585381	05/01/1997	UWC+T		Western	Bridgwater	MLN1
-2311	16/01/1997	AOCL				
742711	05/02/1997	FP		London North Eastern	Mill Lane LC	PBS3
-1607	23/03/1997	AHB				
675095	09/04/1997	MCB			Craven Arms	SHL0
684810	11/05/1997	FP		London North Western	Lime Kiln	CMD2
-1608	25/06/1997	UWC+T				
506979	02/07/1997	AHB		London North Eastern	Burton Lane No.2	GRS2
580583	29/08/1997	UWC		South East	Mellish House	SUD0
694535	05/09/1997	FP		London North Eastern	Sportsfield LC	WAG1
-1609	12/10/1997	UWC+T				
746438	18/10/1997	FP		South East	Newtown LC	XTD0
746463	24/10/1997	FP		South East	Peas Marsh LC	WPH1
771004	31/10/1997	FP		South East	Green Lane LC	BTH3
750212	12/12/1997	FP		London North Western	Hard Platts No 1 LC	GJC0
759316	20/01/1998	FP		London North Eastern	Manthorpe LC	ECM1
766071	17/02/1998	MCB+CCTV		London North Eastern	Rossington L C	ECM1
767172	20/02/1998	UWC+T				
788651	12/05/1998	FP		London North Western	Clifton Country Park LC	MVE1
825831	13/08/1998	UWC		South East	Trinity Lane LC	BGK0
826324	16/08/1998	FP		South East	Dearleap Foot LC	BML2
844966	26/10/1998	FP		London North Eastern	Alexander LC	TJG1
849322	14/11/1998	UWC		London North Western	Woodside Farm LC	DSE0
853422	02/12/1998	UWC+T		South East	Gipsy Lane LC	LTN1
877031	21/03/1999	AHB		London North Western	Hixon LC	CMD2
878634	26/03/1999	UWC+T		London North Eastern	Allsops Lane UWGT LC	SPC5

Event ID	Event date	Crossing type	Level crossing name	Territory description	Location description	ELR code
880696	03/04/1999	UWC+T		London North Eastern	Shady Lane UWGT LC	TJC3
905091	04/07/1999	FP		South East	Eastlands FP LC	XTD0
926659	24/09/1999	FP		London North Eastern	Outwoods FP LC	WNS0
964300	29/02/2000	UWC		South East	Red Cross Lane Level Crossing	BGK0
966587	11/03/2000	UWC+T		London North Eastern	Uttoxeter	NSS0
968377	20/03/2000	FP		London North Eastern	Ferryboat Lane Crossing	PED5
985755	18/05/2000	UWC MWL		London North Western	Norton Level Crossing	CGJ2
992698	02/06/2000	FP		South East	Rochford	SSV0
992763	03/06/2000	FP		South East	Lankesters Foot Crossing Stowmarket	LTN1
995975	22/06/2000	AHB		London North Western	Hixon LC	CMD2
1026379	11/10/2000	FP		Western	SWINDON (South Leaze)	MLN1
1027465	16/10/2000	FP		Scotland	Back Laurencekirk Settlement accommodation crossing	ECN5
1047345	05/01/2001	FP MWL		South East	FISHBOURNE	TBH2
1057639	14/02/2001	UWC		South East	Lancing	BLI1
1063978	08/03/2001	SP			Gwersyllt	WDB1
1066759	20/03/2001	UWC		South East	Bishopstone	STS0
1088247	27/05/2001	FP		Western	Lower Howsell Foot Crossing Newlands East	WAH0
1097949	02/07/2001	AHB		Scotland	Markle AHB LC	ECM8
1102548	14/07/2001	UWC+T		London North Eastern	Boat House LC	NEC2
1105793	22/07/2001	AHB		South East	Pevensey Sluice AHB LC	WJB0
1119653	10/09/2001	MCB+CCTV		London North Western	Maghull	SJO2
1136186	06/11/2001	FP		London North Western	Whitebridge LC Stone Station	CMD2
1147935	23/12/2001	FP		London North Western	Saddleworth LC	MVL3
1147751	23/12/2001	AHB		South East	Sawbridgeworth AHB LC	BGK0
1168257	09/03/2002	UWC+T			Downs Farm	LLA0
1200372	19/06/2002	UWC		London North Western	Bushbury Jcn	RBS2
1204169	01/07/2002	FP		London North Eastern	Cottingham	HBS0
1215742	06/08/2002	FP			Tywyn station crossing	DJP0
1223029	29/08/2002	FP MWL		South East	Johnson's foot crossing	BGK0
1223732	31/08/2002	FP		London North Eastern	Dudley (ECML)	ECM7
1227000	11/09/2002	AHB		South East	Moreton (Dorset)	BML2
1262716	27/11/2002	UWC+T			Heol-y-Deliaid foot crossing	OVE0
1279556	28/12/2002	FP		London North Eastern	Garforth	HUL4
1417471	07/04/2003	AHB		Western	Dunhampstead	BAG2

Event ID	Event date	Crossing type	Level crossing name	Territory description	Location description	ELR code
1478733	20/05/2003	FP		Scotland	Banefield Foot crossing	NEM7
1478787	20/05/2003	FP		South East	Nr Essex Road CCTV lx	BGK0
1502077	06/06/2003	FP		London North Eastern	Thorne North	TJG1
1497873	07/06/2003	FP		London North Western	Rylands FP LC	MVE2
1567416	10/08/2003	FP		London North Eastern	Green Drift crossing	SBR0
1580675	23/08/2003	FP		London North Eastern	Knottingley	WAG1
1626094	06/10/2003	AHB		London North Eastern	Snelland AHB lc	NOB3
1626318	11/10/2003	FP		London North Eastern	Outwoods foot level crossing	WNS0
1684300	08/12/2003	FP		London North Western	Hard Platts No. 1 Foot crossing	GJC0
1693113	15/01/2004	MCB+CCTV		London North Western	Canley	RBS1
1697360	31/01/2004	MCB+CCTV		South East	St Denys	SDP1
1703501	04/02/2004	FP		London North Eastern	Pelham Street	SPD3
1703157	18/02/2004	FP		South East	Baileys Drove Foot Crossing, Wool	BML2
1723879	23/04/2004	FP			Tir-Phil	CAR0
1735445	24/05/2004	UWC	Bostocks	London North Eastern	Upper Leigh LC	NSS
1736888	05/06/2004	AHB		South East	Pevensey Sluice AHB lc	WJB0
1760443	21/08/2004	SP		South East	Gomshall station	RSJ0
1777708	21/10/2004	FP MWL		South East	Fishbourne LC (Main Line)	TBH2
1781781	06/11/2004	AHB		Western	Ufton	BHL
1782388	08/11/2004	FP		London North Eastern	Shepherds Farm crossing	SSK1
1814132	14/03/2005	AHB		Scotland	Kirknewton	ECA2
1816572	23/03/2005	AHB		London North Eastern	Clara Vale AHB LC	NEC2
1819375	03/04/2005	UWC+T		South East	Leigh on Sea	FSS2
1833212	20/05/2005	FP		London North Eastern	Bardon Hill	KSL0
1847708	06/07/2005	AHB		South East	Eastrea AHB LC	EMP0
1850491	18/07/2005	FP		London North Western	Coleshill FC LC	NWO0
1851397	20/07/2005	AHB		London North Eastern	Tilford Road AHB LC	RAC0
1856088	07/08/2005	UWC		South East	March	EMP
1858491	09/08/2005	FP		Western	Fernhill	ABD
1878320	29/10/2005	AHB		London North Eastern	Aslockton AHB L.C	NOG1
1882321	13/11/2005	AHB		South East	Swainsthorpe LC	LTN1
1884262	21/11/2005	FP		London North Eastern	foot crossing near Attenborough station	TSN1
1887480	03/12/2005	SP MWL		South East	Elsenham station Foot Crossing	BGK0
1894650	08/01/2006	FP		London North Eastern	Cattal	HAY1

Event ID	Event date	Crossing type	Level crossing name	Territory description	Location description	ELR code
1899725	26/01/2006	AHB		South East	Wilmington LC	KJE3
1913258	17/03/2006	FP		South East	Pegamoid Foot crossing	BGK0
1926280	03/05/2006	AOCL		London North Eastern	Seacroft AOCL LC	GRS4
1934341	03/06/2006	UWC		London North Eastern	Knottingley	WAG1
1936454	09/06/2006	UWC		Western	Midgham	BHL0
1954808	18/08/2006	FP			Abergavenny	HNL1
1959758	06/09/2006	MCB+CCTV		London North Eastern	Lincoln Street	RAC0
1975185	13/11/2006	FP		London North Eastern	Garforth Moor foot level crossing	HUL4
1986197	01/01/2007	FP		South East	Paggett's footpath level crossing	BLI1
1988721	14/01/2007	FP		London North Eastern	Royston	SBR0
1988502	15/01/2007	FP			Johnstown foot crossing	WSJ2
1991073	24/01/2007	MCB		South East	Wokingham LC	RDG1
1991036	25/01/2007	FP MWL		London North Eastern	Ballast Hole MWL FP crossing	NOB1
1997403	23/02/2007	AHB		Scotland	Gailes LC	AYR4
2036604	31/07/2007	UWC+T		London North Western	Sandringham Avenue LC	CWK3
2039913	11/08/2007	MCB+CCTV		South East	Highams Park	CJC0
2040874	16/08/2007	MCB+CCTV		South East	Horsham Road LC	TBH1
2054226	17/10/2007	UWC+T		London North Western	Penketh Hall LC	SDJ2
2059728	12/11/2007	FP		London North Eastern	Green Drift FP MWL	SBR0
2061463	20/11/2007	FP MWL		Western	Brimscombe footpath crossing	SWM1
2073360	21/01/2008	UWC+T		London North Western	Melrose Avenue UWC-T	CWK3
2073605	22/01/2008	UWC+T		London North Eastern	West Lodge UWC-T	NEC2
2078775	13/02/2008	FP		South East	Leys Lane footpath crossing	ETN0
2088686	27/03/2008	MCB+CCTV		South East	Hythe CCTV LC	COC0
2089464	31/03/2008	SP		Western	Tackley UWG-T LC	DCL0
2093407	16/04/2008	FP		South East	Moor Lane FP LC	SWE0
2099544	10/05/2008	FP MWL		Western	Marsh Barton footpath crossing MWL	DAC0
2099227	10/05/2008	AHB		South East	Blackboy Lane AHB LC	TBH2
2099737	10/05/2008	AHB		London North Eastern	Marston on Dove LC	NSS
2108462	13/06/2008	MCB+CCTV			Pencoed MBC LC	SWM2
2117825	21/07/2008	FP		South East	Itchingfield footpath crossing	TBH1
2145655	22/11/2008	FP		London North Eastern	Bayles & Wylies FP MWL crossing	RAC0
2147218	01/12/2008	FP		London North Eastern	Snuff Mill Lane	HBS0
2150779	17/12/2008	MCB+CCTV		South East	Rainham LC	TLL0

Event ID	Event date	Crossing type	Level crossing name	Territory description	Location description	ELR code
2158325	23/01/2009	AHB		Scotland	Gatehead	BAK0
2186158	07/02/2009	UWC MWL		Scotland	Moulinearn	HGL2
2247464	02/04/2009	UWC+T		London North Eastern	Peth Lane, Ryton, Tyne & Wear (Newcastle-Carlisle line)	NEC2
2248324	03/04/2009	AHB			Eyton	WSJ2
2273424	06/05/2009	FP		Western	Fairfield FP LC	BHL0
2296664	19/05/2009	MCB+CCTV		South East	Horsham Road LC, Crawley	TBH1
2301107	23/05/2009	UWC+T		Western	Trowbridge	BFB0
2381546	12/08/2009	UWC+T		South East	Keysworth LC	BML2
2400966	05/09/2009	FP		London North Eastern	Bridlington	HBS0
2402523	07/09/2009	FP		London North Eastern	Fox Covert foot crossing	WEB0
2468265	02/11/2009	FP		London North Eastern	Attenborough nature reserve	TSN1
2498009	29/12/2009	FP		Western	Chawson	OWW0
2505907	15/01/2010	MCB+CCTV		South East	Horsham Road LC (CCTV)	TBH1
2548930	06/03/2010	AHB		South East	Waterloo (Wokingham) LC	RDG1
2583668	16/05/2010	FP		London North Eastern	Old Gashouse LC (Morley)	MDL1
2602189	14/07/2010	FP		South East	Sherrington LC (Warminster)	SAL0
2608071	09/08/2010	MCB+CCTV		South East	Enfield Lock LC (Ordance Road)	BGK0
2635201	15/11/2010	FP		London North Eastern	Branston GF	DBP1
2641279	11/12/2010	UWC+T		London North Western	Fishermans Path, Freshfield, Formby, Sefton, Merseyside	HXS3
2647179	13/01/2011	UWC		London North Eastern	Cleethorpes	MAC3
2648534	19/01/2011	MCB+CCTV		Western	Morris Hill CCTV LC, Cheltenham Spa, Gloucestershire	BAG2
2650174	29/01/2011	FP		Western	Langport	CCL0
2654008	14/02/2011	FP		South East	Sharpenhurst No.3 LC (Christ's Hospital), West Sussex	TBH1
2679191	13/05/2011	FP MWL		London North Eastern	Norton-on-Tees	STF0
2696545	25/07/2011	AHB		South East	Church Street LC	LTN1
2703867	24/08/2011	FP		South East	Gipsy Lane FP LC, Needham Market, Suffolk	LTN1
2713153	03/10/2011	FP		Western	Mexico Foot Path crossing, Long Rock, Penzance	MLN4
2737800	28/01/2012	FP MWL		South East	Johnsons R/G Footpath crossing, Bishops Stortford, Hertfordshire (Harlow-Audley End line)	BGK0
2740244	08/02/2012	FP		South East	Wool West LC	BML2
2756977	23/04/2012	UWC MWL		London North Eastern	Ivy Farm UWCT, Royston, Hertfordshire (Hitchin-Cambridge line)	SBR0
2759369	02/05/2012	FP		South East	Hoo Hall crossing, Kelvedon, Essex	LTN1
2759419	02/05/2012	FP		London North Eastern	Kings Mill No.1 bridleway crossing, Mansfield, Nottinghamshire	PBS2

Event ID	Event date	Crossing type	Level crossing name	Territory description	Location description	ELR code
2763852	22/05/2012	AHB	Ufton (Aldermaston)	Western	Ufton	BHL
2779902	29/07/2012	AHB		South East	New Fishbourne LC	TBH2
2816076	08/01/2013	FP		London North Eastern	Thorne South	DOW0
2819113	22/01/2013	AHB			Newcastle Rd LC	SYC0
2819688	24/01/2013	FP MWL		South East	Motts Lane (R/G-X) LC (between Witham and Kelvedon)	LTN1
2836320	07/04/2013	UWC+T		London North Eastern	Blackhills Farm UWC	LEN3
2857260	27/06/2013	FP		Western	Mill Lane FP crossing, Marlow	MWB0
2860705	11/07/2013	FP MWL		Western	Springfield Road footpath crossing	MLN1
2880542	03/10/2013	UWC MWL		South East	Dernford R/G UWG crossing	BGK0
2885436	26/10/2013	FP		London North Eastern	Attenborough	TSN1
2890004	12/11/2013	AHB		Western	Sandy Lane LC	DCL
2904257	15/01/2014	FP MWL		London North Eastern	KETTON MCB CROSSING,STAMFORD	PMJ0
2920088	24/03/2014	FP		South East	Cattishall footpath crossing	CCH0
2936828	31/05/2014	AHB		South East	Wharf Road AHB LC	BGK0
2943212	26/06/2014	AHB		South East	Wharf Road AHB LC	BGK0
2957922	27/08/2014	UWC+T		London North Western	Fishermans Path (UWGT)	HXS3
2959549	03/09/2014	FP		South East	Dibleys Foot Crossing	ACR0
2960884	09/09/2014	FP		South East	Clappers Lane foot crossing	BLI1
2962496	16/09/2014	FP		London North Eastern	Wyke, Lightcliffe golf course footpath crossing	MRB0
2973711	04/11/2014	AHB		South East	Sandhill AHB LC	BGK0
2977323	19/11/2014	YDS			Gretsy Lane No.1	
2982776	13/12/2014	FP		London North Eastern	Hipperholme (AKA Shibeden Park) footpath level crossing	MRB0
2993967	08/02/2015	FP		South East	Glebe Way FP LC	VIR0
2994495	09/02/2015	FP		London North Western	Nelson	GJC0
3007093	08/04/2015	FP MWL		South East	Cannons Mills LC	BGK0
3060084	24/11/2015	FP		London North Western	Old Stoke Road LC	PRA0
3077564	15/02/2016	UWC+T		South East	Tide Mills UWC	STS0
3079518	23/02/2016	FP		South East	No.22 Grimston Lane Footpath Crossing	FEL0
3080324	27/02/2016	MCB+CCTV		South East	Shoreham Station CCTV, Buckingham Road	BLI1
3114917	23/07/2016	MCB+CCTV		South East	Stockbridge Road LC	TBH2
3128510	19/09/2016	MCB+CCTV		South East	Mount Pleasant	BML1
3132202	05/10/2016	FP		South East	Alice Holt foot crossing	PAA2
3140032	09/11/2016	FP		London North Western	Old Stoke Road Public FP Crossing	PRA0

Event ID	Event date	Crossing type	Level crossing name	Territory description	Location description	ELR code
3147271	09/12/2016	FP		London North Western	Brierfield	GJC0
-27299	06/03/2017	UWC			Stokeswood UWC	
-27741	24/03/2017	footpath			Nowhere foot crossing	
-28641	25/03/2017	footpath			Starcross Level Crossing	
-27655	15/05/2017	AHB			Dimmocks Cote AHB	
-27679	17/05/2017	footpath	Nature Reserve		Beeston	
-27725	01/06/2017	UWC			Trenos	
-28163	26/09/2017	footpath			Wallows Lane FP	
-28589	30/01/2018	MCB-CCTV			Lincoln Street	
-28679	17/02/2018	AHB			Barns Green AHB Level Crossing	

Appendix B: Python code used to run machine learning tests

This appendix provides the Python source code that was used to perform the machine learning tests. The same source code was used for all tests: configuration of the code is required to select the features and tests to be performed. The code relies on software libraries being installed, specifically `datetime`, `numpy` and `sklearn`.

```
# User definitions ==============================================================
#   The path to the source data
file_path      = r"C:\Users\Staff\ML Analysis"
file_name      = r"\CALC 12 for import.txt"
type_column    = 2

# The list of features and the label for this analysis
operational    = [3, 4, 5, 6, 7, 8, 9]
hazards        = [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22]
controls       = [23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36]
all            = operational + hazards + controls
features       = [9]

# The list of types of level crossings for this analysis
# Passive Vehicular Public:                         5
# Passive Vehicular Private or Staff:               6
# Automatic Vehicular Public:                       7
# Automatic Vehicular Private or Staff:             8
# Railway-controlled Vehicular Public:              9
# Railway-controlled Vehicular Private or Staff:   10
types          = [10]

# Label: Whether there are any observed collisions [0, 1] : 40
label          = 40

testing_proportion  = 0.2
number_of_test_runs = 1000
discriminator_total = 0

# Start code ====================================================================
import datetime
print(datetime.datetime.now(), "> Importing libraries")
import numpy
import sklearn
from   sklearn.tree           import DecisionTreeClassifier, plot_tree
from   sklearn.ensemble       import RandomForestClassifier
from   sklearn.neural_network import MLPClassifier
from   sklearn.svm            import SVC

# Get the source data and format it into X and y data structures _________________
print(datetime.datetime.now(), "> Importing source data")
source_data = numpy.genfromtxt(file_path+file_name, delimiter="\t")
no_of_rows  = len(source_data)
no_of_cols  = len(source_data[0])

# Using the convention: X is the array of features, y is the  vector of labels
X = []
y = []
```

```
# Extract the correct columns of the source data into the X and y data structures
for row in range(1, no_of_rows):               # Literal 1 at start of range because the
first row (zero) contains headers
    if source_data[row][type_column] in types:  # collect data only if it is in the list
of types to collect
        data_collector = []                    # a temporary array to collect values for
each row in X
        for column in range(no_of_cols):
            if column in features:             # if the current column number is in the
list of feature numbers we're selecting
                data_collector.append(source_data[row][column])
        X.append(data_collector)
        y.append(source_data[row][label])
# Post condition: the data are now in X and y structures per the features and label
values

# Run the tests =====================================================================

for test_count in range(number_of_test_runs):
    print("=========================================================================")
    print(datetime.datetime.now(), "> Running model ", test_count + 1, " ________ ")

# ============================== DEFINE THE MODEL BELOW ================================
    test_name = "Decision tree (sklearn.tree.DecisionTreeClassifier)"
    model = DecisionTreeClassifier()

    #test_name = "Random forest (sklearn.ensemble.RandomForestClassifier)"
    #model = RandomForestClassifier(max_depth=5, n_estimators=10)

    #test_name = "ANN MLP (sklearn.neural_network.MLPClassifier)"
    #model = MLPClassifier(alpha=1, max_iter=1000)

    #test_name = "RBF SVM (sklearn.svm.SVC)"
    #model = SVC(gamma=2, C=1)
# ======================================================================================

    x_train, x_test, y_train, y_test = sklearn.model_selection.train_test_split(X, y,
test_size = testing_proportion)
    model.fit(x_train, y_train)
    predicted = model.predict(x_test)

    # Get the results ==================================================================
    true_positive = 0
    true_negative = 0
    false_positive = 0
    false_negative = 0

    for i in range(len(predicted)):
        if predicted[i] == 1 and y_test[i] == 1:
            true_positive +=1
        if predicted[i] == 0 and y_test[i] == 0:
            true_negative +=1
        if predicted[i] == 1 and y_test[i] == 0:
            false_positive += 1
        if predicted[i] == 0 and y_test[i] == 1:
            false_negative +=1

    '''print(datetime.datetime.now(), "> Calculating results")
    print("Test: ", test_name)
    print("Total rows in source data           : ", no_of_rows)
    print("Number of rows imported             : ", len(X))
    print("Test set size                       : ", testing_proportion)
    print("Expected number of rows in test set : ", len(X) * testing_proportion)
    print("Actual   number of rows in test set : ", len(predicted))
    print()
    print("Types          : ", types)
    print("Features       : ", features)
    print("Label          : ", label)
    print()
    print("True  positive : ", true_positive)
    print("True  negative : ", true_negative)
```

```
    print("False positive : ", false_positive)
    print("False negative : ", false_negative)
    print()
    print("Results check  (TP + TN + FP + FN) :", true_positive + true_negative +
false_positive + false_negative)
    print()'''

    naive_accuracy = 0 if true_positive == 0 else (true_positive + true_negative) /
len(predicted)
    positive_discriminator = 0 if true_positive == 0 else true_positive / (true_positive
+ false_negative)

    print("Naive accuracy         (TP + TN) / (total)  : ", naive_accuracy)
    print("Positive discriminator  TP / ( TP + FN )    : ", positive_discriminator)
    discriminator_total += positive_discriminator
    print("=======================================================================\n")

average_discriminator = 0 if discriminator_total == 0 else discriminator_total /
number_of_test_runs #Sometimes there are no true positives to match, so we get a DBZ
error

print("\n ------------ TEST SUMMARY ------------")
print("Test\tTypes of level crossing\tFeatures\tLabel\tNumber of tests\tAverage
discriminator value\n")
print("%s\t%s\t%s\t%s\t%d\t%f" %(test_name, types, features, label, number_of_test_runs,
average_discriminator))
print(" --------------------------------------")
print(datetime.datetime.now(), "> Test end")
```